Reject, Resample, Repeat: Understanding Parallel Reasoning in Language Model Inference

Noah Golowich
UT Austin
nzg@cs.utexas.edu

Fan Chen
MIT
fanchen@mit.edu

Dhruv Rohatgi
UC Berkeley
drohatgi@mit.edu

Raghav Singhal
NYU
singhal.raghav@gmail.com

Carles Domingo-Enrich
Microsoft Research
carlesd@microsoft.com

Dylan J. Foster
Microsoft Research
dylanfoster@microsoft.com

Akshay Krishnamurthy
Microsoft Research
akshaykr@microsoft.com

Abstract

Inference-time methods that generate, aggregate, and prune multiple parallel reasoning traces have emerged as a powerful paradigm for steering large language models, yet we lack a principled understanding of their accuracy–cost tradeoffs. We develop such an understanding for *multi-particle methods*, which maintain multiple partial chains of thought and use a *process reward model (PRM)* to adaptively score, prune, and replicate them. We focus on *Sequential Monte Carlo (SMC)*, the simplest and most canonical such method. Despite SMC's long history in statistics and recent adoption for steering language and diffusion models, non-asymptotic guarantees have remained elusive. We provide (1) a simple, user-friendly analysis identifying two natural criteria which suffice for non-asymptotic convergence of SMC; (2) a simple modification of SMC achieving stronger, horizon-free guarantees when the PRM is near-perfect; and (3) a fundamental limit faced by *all* myopic multi-particle methods in the presence of PRM approximation errors. Empirically, we find that our criteria effectively predict the sampling error of SMC though not necessarily its final accuracy, paving the way for future work incorporating theoretical perspectives beyond sampling.

1 Introduction

Recent work on large language models (LLMs) has demonstrated the value of *inference-time algorithms* that *steer* models towards higher-quality outputs. For instance, existing work has experimented with a combination of using *parallel* generations—such as multiple chains of thought—from LLMs [3, 35], techniques to *aggregate* such generations [16, 53, 63], and ways to *prune or compress* them to fit in the context window for subsequent model calls [56, 59]. Even with *no additional training*, such interventions can significantly improve performance on challenging tasks such as mathematical reasoning and question answering. However, these methods are largely ad hoc, since we lack a unifying theoretical framework that can *explain* the benefits of different inference-time algorithms and *guide* algorithm design towards the most effective ones. In this paper, we ask:

Can we build a rigorous theory for inference methods that adaptively steer the generation process and characterize how accurately they sample from a target distribution?

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

To do so, we study the paradigm of *guided LLM generation with a process reward model* [42, 50], where the process reward model numerically scores partial generations. In this paradigm, aggregation and pruning are naturally implemented by *multi-particle* algorithms, including the popular *Sequential Monte Carlo (SMC)* method, which use the process reward model to adaptively prune and replicate multiple *particles* (namely, partial generations from the base model), as tokens are generated. These strategies can have robust empirical benefits over simpler approaches such as Best-of-N; see e.g. Figure 2 or Puri et al. [42].

SMC is perhaps the simplest and most canonical multi-particle method, with a long history in statistics and signal processing—where multi-particle methods are known as *particle filtering* methods [1, 4, 8, 10, 12, 13, 21, 37, 40]. More recently, SMC and closely related methods have seen widespread adoption for steering LLMs [14, 22, 31, 33, 34, 41, 55, 62] and diffusion models [47, 48]. Despite this broad applicability, *non-asymptotic guarantees* have remained elusive. In this paper, we address this challenge by focusing on SMC, providing a simple, user-friendly analysis and identifying benefits and limitations of more sophisticated multi-particle methods. As inference-time algorithms grow increasingly complex, utilizing models not just for generation and scoring, but also for operations such as splicing [16] or summarizing [60] generations, we view our analysis as a necessary first step towards a principled understanding of such methods.

1.1 Setting: Guided Generation with Imperfect Process Rewards

For a fixed prompt, a language model defines a distribution $\pi_{\mathtt{ref}} \in \Delta(\mathcal{A}^H)$ over sequences of H *actions* in some set $\mathcal{A}$.[1] One can think of $\mathcal{A}$ as the space of *tokens*, or more generally as blocks of tokens. Given a sequence $a_{1:h}$, we can efficiently sample from $\pi_{\mathtt{ref}}(a_{h+1} = \cdot \mid a_{1:h})$, which enables autoregressive sampling from $\pi_{\mathtt{ref}}$.

We model the task of *steering* the language model towards some desired reward as a *sampling* problem, in line with recent work on LLM inference and post-training [15, 19, 43, 57].[2] Given query access to a reward function $r^\star : \mathcal{A}^H \to \mathbb{R}_{\geq 0}$, we aim to sample from the *target distribution* $\pi^\star_H(a_{1:H}) \propto \pi_{\mathtt{ref}}(a_{1:H}) r^\star(a_{1:H})$, which upweights high-reward outputs relative to the base model. For instance, in the context of mathematical reasoning, $r^\star(a_{1:H}) \in \{0, 1\}$ measures whether $a_{1:H}$ correctly solves the prompt.

Imperfect process rewards. We assume access to a *process reward model (PRM)*, which is a learned scoring function that evaluates incomplete generations, predicting, for example, how likely a partial chain of thought is to ultimately arrive at a correct answer. Following [43], we formalize the PRM as an *approximate value function*: given a sequence $a_{1:h}$ of any length, the PRM provides an estimate $\widehat{V}(a_{1:h}) \in \mathbb{R}_{\geq 0}$ of the *expected reward* of $a_{1:h}$ under $\pi_{\mathtt{ref}}$, i.e., $V^\star(a_{1:h}) := \mathbb{E}_{a'_{1:H} \sim \pi_{\mathtt{ref}}}[r^\star(a'_{1:H}) \mid a'_{1:h} = a_{1:h}]$. Since we are given query access to $r^\star$, it is without loss of generality to assume that $V^\star(a_{1:H}) = \widehat{V}(a_{1:H})$ (as we will do).

In practice, there has been extensive effort to train transformers specifically to be PRMs [32, 51, 52, 61] for certain problems, such as math reasoning. If such a fine-tuned PRM adapted to a desired reward function $r^\star$ is not available, there are several other popular choices for $\widehat{V}$—e.g. LLM-as-a-judge [49] or raw LLM probabilities [28, 31]; see also Section 4.

The key algorithmic challenge in this setting is that the PRM is imperfect: errors at individual generation steps can compound across the sequence, so naïve approaches suffer from error amplification that grows with the sequence length H [43]. To address this challenge, numerous parallel and sequential approaches have been proposed, including SMC [42], block rejection sampling [38], and backtracking [2]. Thus far, provable guarantees are only known for backtracking [43]—an inherently sequential approach. We ask: *given access to $\pi_{\mathtt{ref}}$ and $\widehat{V}$ as above, what structural properties enable approximate sampling from $\pi^\star_H$* ***via efficient multi-particle methods****?*

[1] For simplicity, we assume that the language model produces generations of fixed length. Variable-length generations with a known maximum length can be handled by padding with an absorbing end-of-sequence token.

[2] See also the concurrent and independent work [64].

1.2 Theory: Principled Analysis of Multi-Particle Methods for LM Inference

Contribution I: Simple criteria for success of SMC. Our first result identifies two **key properties** that enable the success of SMC. For $h \in [H]$, let $\pi_h^\star, \widehat{\pi}_h \in \Delta(\mathcal{A}^h)$ denote the *true* and *approximate intermediate target distributions*, defined by $\pi_h^\star(a_{1:h}) \propto \pi_{\mathtt{ref}}(a_{1:h})V^\star(a_{1:h})$ and $\widehat{\pi}_h(a_{1:h}) \propto \pi_{\mathtt{ref}}(a_{1:h})\widehat{V}(a_{1:h})$. Then we show the following:

Theorem 1.1 (Corollary of Theorem 3.2)**.** *Suppose that the following hold.*
(1) *Bounded* action-level coverage*: for all h, $a_{1:h+1}$, $\pi^\star(a_{h+1} \mid a_{1:h})/\pi_{\mathtt{ref}}(a_{h+1} \mid a_{1:h}) \le C_{\mathsf{act}}$.*
(2) *Bounded* χ^2-divergences *(bounding the error of $\widehat{V}$ with respect to $V^\star$): $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) \le C_{\chi^2}$.*

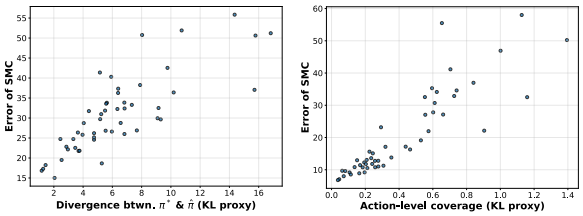
Then SMC with N particles samples from a distribution $\widehat{\mu}$ satisfying $D_{\mathsf{TV}}(\widehat{\mu}, \pi^\star) \le \sqrt{\frac{H^2 C_{\mathsf{act}}(C_{\chi^2}+1)}{N}}$.

This bound strengthens the guarantees for the backtracking-based algorithm VGB of Rohatgi et al. [43]. Moreover, the parallel runtime of SMC is $O(H)$ whereas VGB uses time $\Omega(H^2)$ (and is inherently sequential). See Appendix B for additional discussion.

Remark 1.2 (Connection to literature on refinements of SMC)**.** *Interestingly, the two **key properties** listed above map quite cleanly onto various efforts in prior work focused on refining SMC algorithms. The latter of the two conditions (that $\widehat{V}$ approximates $V^\star$) is related to the line of work on* twisting targets *for SMC. In particular, the distributions $\widehat{\pi}_h$ introduced above are often referred to as (intermediate) target distributions, and prior work has explored how to choose $\widehat{V}$ so as to bring $\widehat{\pi}_h$ as close as possible to $\pi_h^\star$ for all $h \in [H]$ [23, 25, 29, 62], which is a natural goal in light of Theorem 1.1.*

Moreover, the first condition (on action-level coverage) is related to the extensive literature on changing the proposal distribution $\pi_{\mathtt{ref},h}$ to better approximate the locally optimal proposal distribution *[13, 39], which is defined as the distribution $\widehat{\pi}_{h+1}(a_{h+1} = \cdot \mid a_{1:h})$. In the event that $\widehat{\pi}_{h+1} \approx \pi_{h+1}^\star$, then the locally optimal proposal distribution has action-level coverage close to 1, the smallest possible value. Thus, as long as $\widehat{V}$ is sufficiently accurate, modifying $\pi_{\mathtt{ref},h}$ so as to approximate the locally optimal proposal is a way to decrease action-level coverage. Summarizing,* ***our result helps to unify the existing literature on variants of SMC by showing that they are effectively minimizing the key quantities in Theorem 1.1.***

Contribution II: Beyond Sequential Monte Carlo. Can multi-particle methods beyond SMC achieve stronger guarantees? First, we show that if $\widehat{\pi}_h$ is close to $\pi_h^\star$ in a stronger L_∞ sense, then it is possible to achieve *exponential decay* in sampling error (rather than polynomial decay, as in Theorem 1.1) by wrapping SMC in an outer rejection sampling loop (Theorem 3.6). Second, we observe that even when the PRM is *perfect*, and hence $C_{\chi^2} = 0$, SMC requires $\Omega(\sqrt{H})$ particles to achieve non-trivial sampling accuracy (Proposition D.8), even though there is a simple one-particle algorithm for this special case [43]. We show that a simple modification to SMC—*Sequential Monte Carlo with Rejection Sampling (SMC-RS, Algorithm 2)*—avoids this pathology, recovering Theorem 1.1 as well as the one-particle guarantee (Theorem 3.8).



(a) $D_{\mathsf{KL}}(\pi_h^\star \,\|\, \widehat{\pi}_h)$. (b) Action-level coverage.

Figure 1: Empirical validation for our theory on the *prompt-switching task* (see Section 4). (a) We fix $\pi_{\mathtt{ref}} = \pi^\star$ (thus fixing action-level coverage) and vary $\widehat{V}$; observe that $D_{\mathsf{KL}}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ predicts the sampling error of SMC. (b) We vary action-level coverage across prompts while keeping $\widehat{V} = V^\star$ fixed; observe that action-level coverage predicts SMC's sampling error.

Contribution III: Limits of multi-particle methods. A key limitation of SMC (and our modifications) is that even when $\pi_{\mathtt{ref}}$ has bounded action-level coverage (i.e. $C_{\mathsf{act}} = O(1)$) and the approximate value function $\widehat{V}$ is within a constant factor of the truth (and hence $C_{\chi^2} = O(1)$), our algorithm requires maintaining $\mathrm{poly}(H)$ particles and hence requires superlinear work. In Theorem 3.9 we show that fully avoiding this lower bound would require *lookahead*: any *myopic* method needs at least $\Omega(\log H / \log\log H)$ particles (Theorem 3.9).

1.3 Empirics: Does the Theory Predict the Performance of SMC in LMs?

We investigate empirically whether the quantities identified by Theorem 1.1 indeed predict the performance of SMC on LLM sampling problems. First, we consider a simple "prompt switching" task in which the base model $\pi_{\mathtt{ref}}$ corresponds to sampling from an LLM given some prompt $\mathsf{p}_{\mathtt{ref}}$, and $\pi^\star$ corresponds to sampling from that LLM given another prompt $\mathsf{p}^\star$ (Section 4). In Figure 1, we individually control for either the divergence between $\pi_h^\star$ and $\widehat{\pi}_h$, as measured by a KL proxy (Figure 1a) or action-level coverage (Figure 1b), by varying $\mathsf{p}_{\mathtt{ref}}$, $\mathsf{p}^\star$, and $\widehat{V}$. When either quantity is varied and the other is fixed, we see a strong correlation with SMC's sampling error.

Evaluating SMC on math reasoning tasks. Finally, we consider the performance of SMC on challenging math reasoning benchmarks including AIME and Math500. In Figure 2, we plot, for each problem in Math500, the accuracy of SMC with N particles against that of Best-of-N (with $N = 32$); each of these quantities was estimated by running the requisite algorithm 100 times per problem. As most points lie below the line $y = x$, we see that SMC improves performance over Best-of-N in a *uniform sense over all problems*, which extends prior observations in [6, 42] which observed that SMC improves upon Best-of-N in an *on-average* sense over such benchmarks. Despite this strong evidence for superiority of SMC, as we discuss further in Section 4.3, the quality of the PRM $\widehat{V}$ (as measured by an estimate of the chi-square divergence between $\widehat{\pi}_h$ and $\pi_h^\star$) does not seem to positively correlate with performance, as often *larger divergence* leads to *higher accuracy*. ***We leave as an intriguing open question the development of a more refined framework that captures performance on such benchmarks.***

2 Preliminaries

Our theory applies in a setup, discussed below, that generalizes Section 1.1 and captures a broader range of settings including, e.g., guidance for diffusion models. Let $\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_H$ be arbitrary sets, and let $\mathcal{X} := \mathcal{X}_0 \sqcup \mathcal{X}_1 \sqcup \cdots \sqcup \mathcal{X}_H$ be their disjoint union. Let $\pi_{\mathtt{ref}} = (\pi_{\mathtt{ref},h})_{h=0}^{H-1}$ be a collection of transition kernels $\pi_{\mathtt{ref},h} : \mathcal{X}_h \to \Delta(\mathcal{X}_{h+1})$. Without loss of generality, assume that $\mathcal{X}_0 = \{\perp\}$ is a single element. The transition kernels induce a Markov chain $x_0 \to x_1 \to x_2 \cdots \to x_H$, where $x_{h+1} \sim \pi_{\mathtt{ref},h}(\cdot \mid x_h)$ for $0 \le h \le H-1$; we drop the subscript h when clear from context. We write $x_{1:H} \sim \pi_{\mathtt{ref}}$ to denote a sample from the Markov chain. Let $\pi_h \in \Delta(\mathcal{X}_h)$ be the induced marginal distribution of x_h.

Remark 2.1 (Special case: autoregressive generation)**.** *In the autoregressive generation setting of Section 1, $x_h = a_{1:h}$ represents the first h tokens of a chain of thought (so $\mathcal{X}_h = \mathcal{A}^h$), and $\pi_{\mathtt{ref},h}(\cdot \mid a_{1:h})$ represents a pre-trained LM which samples a_{h+1} (given $a_{1:h}$) and appends it to $a_{1:h}$.*

As in the setup of Section 1, given a terminal reward function $r^\star : \mathcal{X}_H \to \mathbb{R}_{\ge 0}$, the *value function* $V^\star : \mathcal{X} \to \mathbb{R}_{\ge 0}$ is defined, for $x \in \mathcal{X}_h$, by $V^\star(x) = \mathbb{E}_{x_{1:H} \sim \pi_{\mathtt{ref}}}[r^\star(x_H) \mid x_h = x]$. Note that $V^\star(x_H) = r^\star(x_H)$ for $x_H \in \mathcal{X}_H$. Here $V^\star(x_h)$ measures the expected quality of a partial generation x_h, e.g., how likely a partial chain of thought is to lead to a correct answer.

For $h \in [H]$, define the distribution $\pi_h^\star \in \Delta(\mathcal{X}_h)$ by $\pi_h^\star(x) \propto \pi_h(x) \cdot V^\star(x)$. One may check that the distributions $\pi_h^\star$ are induced by the Markov chain with kernel $\pi^\star(x_{h+1} \mid x_h) := \frac{\pi_{\mathtt{ref},h}(x_{h+1} \mid x_h) V^\star(x_{h+1})}{V^\star(x_h)}$ using that $\sum_{x_h} \pi_h^\star(x_h)\pi^\star(x_{h+1} \mid x_h) = \pi_{h+1}^\star(x_{h+1})$.[3]

Process reward model. The kernel $\pi^\star$ describes how an ideal generation process (one with access to the true value function $V^\star$) would extend partial generations. Multi-particle methods such as SMC use a PRM $\widehat{V}$ to approximate this ideal process. The PRM is denoted $\widehat{V} : \mathcal{X} \to \mathbb{R}_{\ge 0}$, and satisfies $\widehat{V}(x_H) = r^\star(x_H)$ for $x_H \in \mathcal{X}_H$. We define $\widehat{Z}_h := \mathbb{E}_{x \sim \pi_h}[\widehat{V}(x)]$ for $h \in [H]$, and let $\widehat{\pi}_h \in \Delta(\mathcal{X}_h)$ be given by $\widehat{\pi}_h(x) = \pi_h(x) \cdot \widehat{V}(x)/\widehat{Z}_h$; it is the approximation to $\pi_h^\star$ induced by the PRM $\widehat{V}$.

3 Theoretical Analysis of Multi-Particle Methods

In this section, we develop non-asymptotic guarantees for multi-particle methods.

[3] We note the following slight abuse of notation: we use $\pi^\star$ with no subscript to denote the transition kernel above, and $\pi_h^\star$ with a subscript to denote the corresponding intermediate *distributions*.

Algorithm 1 Sequential Monte Carlo

Input: Transition kernel $\pi_{\mathtt{ref}}$, value function $\widehat{V}$.
Parameter: Number of particles $N \geq 1$.
Initialize $\widehat{\nu}_0 = \delta_\perp$.
for $h = 1, \cdots, H$ **do**
  Sample N particles $\tilde{x}^1_{h-1}, \ldots, \tilde{x}^N_{h-1} \overset{\text{i.i.d.}}{\sim} \widehat{\nu}_{h-1}(\cdot)$.
  Generate $x^i_h \sim \pi_{\mathtt{ref},h-1}(\cdot \mid \tilde{x}^i_{h-1})$ independently for $i \in [N]$.
  Define weights $w^i_h = \frac{\widehat{V}(x^i_h)}{\widehat{V}(\tilde{x}^i_{h-1})}$.
  Define weighted empirical measure $\widehat{\nu}_h := \frac{1}{W_h} \sum_{i=1}^N w^i_h \delta_{x^i_h}$, where $W_h = \sum_{i=1}^N w^i_h$.
Define $\widehat{W}_h := \widehat{V}(\perp) \cdot \prod_{i=1}^h \frac{W_i}{N}$.
Output Option 1: Output $x_H \sim \widehat{\nu}_H$.
Output Option 2: With prob. $\min\left\{\frac{\widehat{W}_H}{2\mathsf{C}_\infty \widehat{V}(\perp)}, 1\right\}$, output $x \sim \widehat{\nu}_H$, else restart the algorithm.

3.1 Non-Asymptotic Convergence of Sequential Monte Carlo

SMC (Algorithm 1) maintains N particles (partial generations) in parallel. At each step, every particle is extended by one step using the base model $\pi_{\mathtt{ref}}$, scored by the PRM $\widehat{V}$, and then the algorithm resamples: particles are drawn with probability proportional to their PRM scores, so that high-scoring particles are replicated and low-scoring ones are discarded. Under what conditions on the base model and PRM does this procedure converge to the target distribution $\pi^\star_H$? We identify two natural criteria.

The first criterion is bounded *action-level coverage*, which requires that the base model $\pi_{\mathtt{ref}}$ assigns reasonable probability to the continuations favored by the target distribution $\pi^\star$.

Assumption 3.1 (Action-level coverage)**.** *There exists parameter $C_{\mathsf{act}} \geq 1$ such that for all $h \geq 0$ and $x_h \in \mathcal{X}_h$, $V^\star(x_{h+1}) \leq C_{\mathsf{act}} V^\star(x_h)$ for any $x_{h+1} \in \mathrm{supp}(\pi_{\mathtt{ref}}(\cdot \mid x_h))$.*

We call this condition *action-level coverage* because $\frac{V^\star(x_{h+1})}{V^\star(x_h)} = \frac{\pi^\star(x_{h+1}|x_h)}{\pi_{\mathtt{ref},h}(x_{h+1}|x_h)}$ is exactly the density ratio between the conditional distributions $\pi^\star$ and $\pi_{\mathtt{ref},h}$.[4]

The second criterion is bounded χ^2-divergences between $\pi^\star_h$ and $\widehat{\pi}_h$ for all h, which controls how accurately the PRM scores partial generations at each step. Quantitatively, the two criteria yield the following guarantee for SMC, which directly implies Theorem 1.1.

Theorem 3.2 (Non-asymptotic convergence for SMC)**.** *Under Assumption 3.1, SMC with N particles achieves $D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq \sqrt{\frac{C_{\mathsf{act}}}{N}}\left(H + \sum_{h=1}^{H-1} \sqrt{D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h)}\right)$.*

Intuition. The sampling error of SMC is controlled by the error of its estimate $\widehat{W}_H$ of the normalizing constant Z, which we decompose into H martingale increments, one per step (Appendix D.2). At each step, resampling replaces an exact expectation by an average over N particles, incurring a fluctuation of order $1/\sqrt{N}$ times the standard deviation of the one-step weights. Action-level coverage bounds how much these weights can vary due to a single step of generation (yielding the factor $\sqrt{C_{\mathsf{act}}}$), while $D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h)$ captures the additional variability due to the mismatch between $\widehat{V}$ and $V^\star$. Summing these errors over the H steps yields Theorem 3.2.

We remark that bounding $D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h)$ only requires *average-case* closeness between $\widehat{V}$ and $V^\star$ (Fact B.2). Existing theory [45, 64] recovers an analogous bound that requires *worst-case* closeness:

Assumption 3.3 (L^∞ approximation quality)**.** *There exists parameter $\mathsf{C}_\infty \geq 1$ such that $\mathsf{C}_\infty^{-1} \leq \frac{V^\star(x)}{\widehat{V}(x)} \leq \mathsf{C}_\infty$ for any $x \in \mathcal{X}$.*

While the proof technique of Schweizer [45] can relax this assumption to some extent, it cannot recover Theorem 3.2 (see Appendix D.3). Moreover, as we show below, we can strengthen Theorem 3.2 further to tolerate *heavy-tailed errors* in $\widehat{V}$.

[4] We remark that Assumption 3.1 can be relaxed to the *average* sense (cf. Theorem D.6).

3.1.1 Robustness to Heavy-Tailed PRM Errors

While the χ^2-divergences in Theorem 3.2 provide an easy-to-interpret upper bound, $D_{\chi^2}(\pi^\star_h \| \widehat{\pi}_h)$ is sensitive to the tail behavior of $\widehat{\pi}_h$ with respect to $\pi^\star_h$ (equivalently, the tail behavior of $\widehat{V}$ relative to $V^\star$); in the case of heavy-tailed errors, it can blow up and render Theorem 3.2 vacuous.

However, the goal of SMC is to produce a sample from an *approximately* correct distribution, so the tail behavior should only appear as an error probability. This motivates the following notion of *coverage* [7], which can be bounded in terms of χ^2-divergence but not vice versa——in particular, it can be arbitrarily small even when $D_{\chi^2}(\pi^\star_h \| \widehat{\pi}_h) = \infty$.

Definition 3.4. *For parameter $M \geq 1$, $h \in [H]$, we define* $\mathrm{Cov}_M(\pi^\star_h \| \widehat{\pi}_h) = \mathbb{P}_{x\sim\pi^\star_h}\left(\frac{\pi^\star_h(x)}{\widehat{\pi}_h(x)} \geq M\right)$.

Theorem 3.5 (SMC under heavy-tailed PRM errors)**.** *Under Assumption 3.1, SMC with N particles achieves the following for any $M \geq 1$:* $D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq H\sqrt{\frac{MC_{\mathsf{act}}}{N}} + \sum_{h=1}^{H-1} \mathrm{Cov}_M(\pi^\star_h \| \widehat{\pi}_h)$.

Let us compare the guarantee of Theorem 3.5 with that of Theorem 3.2: the coverage terms $\mathrm{Cov}_M(\pi^\star_h \| \widehat{\pi}_h)$ in Theorem 3.5 play the role of the chi-square divergences $D_{\chi^2}(\pi^\star_h \| \widehat{\pi}_h)$ in Theorem 3.2, capturing the amount by which $\widehat{\pi}_h$ deviates from $\pi^\star_h$. A key difference, though, is that for any fixed value of M, the TV error between $\mathbb{E}[\widehat{\nu}_H]$ and $\pi^\star_H$ in Theorem 3.5 does not approach 0 as $N \to \infty$, i.e., we pay for the mass $\mathrm{Cov}_M(\pi^\star_h \| \widehat{\pi}_h)$ *additively*.

3.2 Improving SMC: Logarithmic Dependence on Accuracy

Theorem 3.2 requires $N = \mathrm{poly}(1/\delta)$ particles to achieve sampling error δ. Can we do better? Below, we show that a simple modification to SMC (Output Option 2 of Algorithm 1) reduces this to $N = \mathrm{poly}(H) \cdot \log(1/\delta)$ when $\widehat{V}$ and $V^\star$ are close in a worst-case sense (Assumption 3.3).

Theorem 3.6 (Fast-rate SMC under L^∞ condition)**.** *Under Assumptions 3.1 and 3.3, Algorithm 1 with **Output Option 2** samples from a distribution μ that satisfies $D_{\mathsf{TV}}(\mu, \pi^\star_H) \leq \delta$, as long as $N \geq \Omega(H\mathsf{C}^4_\infty C^2_{\mathsf{act}} \log(\mathsf{C}_\infty H/\delta))$, using a number of outer loop steps that is bounded by $O(\mathsf{C}^2_\infty \log(1/\delta))$.*

This qualitatively matches the convergence rate of the backtracking-based method VGB of Rohatgi et al. [43], which requires the same assumptions.

3.3 Improving SMC: Horizon-Free Convergence with Near-Perfect PRM

Even when $\widehat{V} = V^\star$, the upper bound in Theorem 3.2 still scales as $\frac{H}{\sqrt{N}}$. In fact, this is a fundamental limitation of SMC, not an artifact of the analysis: even when $\widehat{V} = V^\star$ and $C_{\mathsf{act}} = 2$, SMC cannot achieve $o(1)$ sampling error without at least $N \geq \Omega(\sqrt{H})$ particles (Proposition D.8).[5] Note that if $\widehat{V} = V^\star$, the one-particle algorithm which simply samples $x_h \sim \widehat{\pi}(x_h \mid x_{h-1})$ succeeds in exactly sampling from $\pi^\star$, so there is indeed hope to improve the upper bound of $H/\sqrt{N}$.

To address this limitation, we consider a simple modification to SMC: *Sequential Monte Carlo with Rejection Sampling* (SMC-RS, Algorithm 2 in the appendix). Slight variants of this algorithm have previously been considered under the name *Diffusion Monte Carlo* [24, 44]. In vanilla SMC, every particle produces exactly one successor, and then the algorithm resamples (normalizes) to maintain N particles; this normalization step introduces interference between particles. In SMC-RS, each particle instead produces a *random number* of successors via rejection sampling: a proposed child x_h from parent x_{h-1} is accepted with probability proportional to the PRM ratio $\widehat{V}(x_h)/\widehat{V}(x_{h-1})$. This eliminates the normalization step, so that regardless of the other particles, the conditional distribution $x_h \mid x_{h-1}$ is $\widehat{\pi}(x_h \mid x_{h-1}) \propto \pi_{\mathtt{ref}}(x_h \mid x_{h-1})\widehat{V}(x_h)$.

Importantly, values of x_h for which $\widehat{V}(x_h)$ is larger are more likely to be accepted in the rejection sampling step and thus included in $\mathcal{S}_h$, and thus Algorithm 2 is *implicitly* implementing the reweighting step of SMC (where *explicit* weights w^i_h are defined). In particular, when $N = 1$, SMC-RS reduces to *action-level rejection sampling* [58], which is an exact sampler when $\widehat{V} = V^\star$ (see Appendix B for the definition). We analyze SMC-RS under the following action-level coverage

[5] We leave the problem of determining the optimal H-dependence as an open question.

condition, which we view as conceptually equivalent to Assumption 3.1 (and formally equivalent, up to a factor of C_∞^2 under the L_∞ condition of Assumption 3.3). Assumption 3.7 is needed in order to ensure that a bounded value of the parameter η in Algorithm 2 suffices.

Assumption 3.7. *There exists $\widehat{C}_{\mathsf{act}} \geq 1$ so that for all $h \geq 0$, $x_h \in \mathcal{X}_h$, $\widehat{V}_{h+1}(x_{h+1}) \leq \widehat{C}_{\mathsf{act}} \widehat{V}_h(x_h)$ for all $x_{h+1} \in \mathrm{supp}(\pi_{\mathtt{ref}}(\cdot \mid x_h))$ and $\mathbb{E}_{x_{h+1} \sim \pi_{\mathtt{ref},h}(\cdot|x_h)}[\widehat{V}_{h+1}(x_{h+1})] \geq \widehat{C}_{\mathsf{act}}^{-1} \widehat{V}_h(x_h)$.*

Theorem 3.8 (Horizon-free convergence with near-perfect PRM)**.** *Under Assumption 3.7, suppose that SMC-RS (Algorithm 2) is instantiated with $\eta \geq \widehat{C}_{\mathsf{act}}$. Then SMC-RS with N particles achieves $D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq \frac{1}{\sqrt{N}} \sum_{h=1}^H \sqrt{D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h)}$, and has expected time complexity $O(NH\eta\widehat{C}_{\mathsf{act}})$.*

In particular, suppose $\widehat{V}$ is near-accurate so that $D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h) \leq \epsilon^2$ for all $h \in [H]$. Then SMC-RS achieves $D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq \frac{\epsilon H}{\sqrt{N}}$. If $\epsilon \leq o(1/H)$, then SMC-RS achieves sampling error $o(1)$ with only $N = O(1)$ particles. This is in sharp contrast to SMC, where by Proposition D.8, $N \geq \Omega(\sqrt{H})$ particles are necessary even when $\widehat{V} = V^\star$ and $C_{\mathsf{act}} = O(1)$. Finally, SMC-RS admits analogues of Theorem 3.5 (heavy-tailed errors) and Theorem 3.6 (fast rates); see Theorems E.3 and E.5.

3.4 Limits of Multi-Particle Methods

A key downside of SMC (and the discussed variants) is that, when the PRM is imperfect, avoiding error amplification requires the number of particles N to grow at least linearly in the horizon H. In this section, we show that some horizon dependence is necessary for all *myopic* multi-particle methods. Informally, a myopic multi-particle method is any algorithm that maintains a set of particles $\mathcal{S}_h$ at each step h, and does not use the PRM data from later steps $k > h$ to determine the set $\mathcal{S}_h$. SMC and SMC-RS are two examples of myopic methods; see Appendix G for a formal definition. We prove that even with mild (constant-factor) imperfections in the PRM, any myopic multi-particle method requires nearly $\Omega(\log H)$ particles to even obtain non-trivial *coverage* of $\pi^\star_H$.

Theorem 3.9 (Lower bound for myopic multi-particle methods)**.** *Let $N(H) := \log(H)/(8\log\log(H))$. There is **no myopic multi-particle method** $\mathtt{Alg}$ with the following guarantee for all sufficiently large H: $\mathtt{Alg}$ uses at most $N(H)$ particles, and under Assumptions 3.1 and 3.3 with $C_{\mathsf{act}} = 2$ and $\mathsf{C}_\infty = e^3$, the output distribution ν of $\mathtt{Alg}$ satisfies $\nu(\mathcal{E}) \geq 1/H$ for all events $\mathcal{E} \subset \mathcal{X}_H$ with $\pi^\star_H(\mathcal{E}) \geq 1/2$.*

4 Experiments in Language Models

We now turn to experiments, investigating whether the two natural criteria from our analysis (action-level coverage and PRM accuracy) predict the performance of SMC for language models.

4.1 Controlled Setting: Prompt-Switching Task

A key challenge in validating our theory is that the quantities of interest (action-level coverage, divergence between $\pi^\star_h$ and $\widehat{\pi}_h$) involve the target distribution $\pi^\star_h$, which is typically not readily available. While prior work (e.g., Zhao et al. [62]) has studied how to evaluate the performance of SMC using the algorithm's estimate of the partition function $Z := \mathbb{E}_{\pi_{\mathtt{ref}}}[V^\star(x_H)]$, such techniques are only guaranteed to provide tight bounds if SMC is already accurate, and do not directly address the challenge of estimating quantities such as action-level coverage. We therefore design a controlled setting where both $\pi_{\mathtt{ref}}$ and $\pi^\star$ are efficiently sampleable, enabling direct measurement of our theoretical quantities and SMC's sampling error.

We consider the *prompt-switching task*, where $\pi_{\mathtt{ref}}$ and $\pi^\star$ correspond to the output distribution of a language model under different prompts. Formally, we fix a language model M, and consider the setting of Section 1 where $\mathcal{A}$ is the token space of M and $\mathcal{X}_h = \mathcal{A}^h$. Each instance of the prompt-switching task is determined by 3 prompts, $(\mathsf{p}_{\mathtt{ref}}, \mathsf{p}^\star, \mathsf{p})$, as follows. Given choices of these prompts, we let $\pi_{\mathtt{ref}}$ be the output distribution of M for the prompt $\mathsf{p}_{\mathtt{ref}}$, and we denote it as

$\pi_{\mathsf{ref}}(\cdot) = \mathsf{M}(\cdot \mid \mathsf{p}_{\mathsf{ref}})$. We then choose the terminal reward function to be $r^\star(a_{1:H}) := \frac{\mathsf{M}(a_{1:H}|\mathsf{p}^\star)}{\mathsf{M}(a_{1:H}|\mathsf{p}_{\mathsf{ref}})}$, which gives that, for any intermediate step h, $V^\star(a_{1:h}) = \mathsf{M}(a_{1:h} \mid \mathsf{p}^\star)/\mathsf{M}(a_{1:h} \mid \mathsf{p}_{\mathsf{ref}})$. It follows that $\pi^\star_h(a_{1:h}) = \mathsf{M}(a_{1:h} \mid \mathsf{p}^\star)$. Since it is tractable to sample from M and compute sequence-level probabilities, sampling from $\pi_{\mathsf{ref}}, \pi^\star$ as well as computing $V^\star$ is tractable. Finally, the prompt p determines $\widehat{V}$ in a way that is discussed in Appendices H.2 and H.3.

Evaluation. To evaluate SMC on the prompt-switching task, we consider the following metric, inspired by more complex "LLM-as-judge" techniques and also the use of log-probabilities to characterize the output distributions of LLMs [20]. Given two distributions ν, ν' over sequences of length H, and a set of prompts $\mathcal{P}$, we define the *$\mathcal{P}$-logprob discrepancy* $\mathrm{LPD}_{\mathcal{P}}(\nu, \nu')$ between ν and ν' as the difference between the mean token-level log-probabilities assigned to sequences sampled from ν and ν' under the prompts in $\mathcal{P}$ (summed over all positions h). Formally, for $\nu, \nu' \in \Delta(\mathcal{A}^H)$

$$\mathrm{LPD}_{\mathcal{P}}(\nu, \nu') := \sum_{\mathsf{p} \in \mathcal{P}} \sum_{h=1}^{H} \left| \mathbb{E}_{a_{1:H} \sim \nu} \left[\log \mathsf{M}(a_h \mid \mathsf{p}, a_{1:h-1}) \right] - \mathbb{E}_{a'_{1:H} \sim \nu'} \left[\log \mathsf{M}(a'_h \mid \mathsf{p}, a'_{1:h-1}) \right] \right|.$$

In our evaluations, we choose $\mathcal{P} = \{\mathsf{p}_{\mathsf{ref}}, \mathsf{p}^\star\}$, which is natural in light of the fact that the two most salient prompts in the prompt switching task are $\mathsf{p}_{\mathsf{ref}}$ and $\mathsf{p}^\star$.

4.2 PRM Accuracy and Action-Level Coverage Predict SMC Performance

We consider $k = 50$ instances of the prompt-switching task as described above: each instance $i \in [k]$ is specified by three prompts $((\mathsf{p}_{\mathsf{ref}})^{(i)}, (\mathsf{p}^\star)^{(i)}, \mathsf{p}^{(i)})$. To evaluate the influence of PRM accuracy on SMC performance, we fix $\pi_{\mathsf{ref}} = \pi^\star$ across all instances (in particular, by fixing $\mathsf{p}^{(i)}_{\mathsf{ref}} = (\mathsf{p}^\star)^{(i)}$ *across all* i), thus fixing action-level coverage, and vary the PRM $\widehat{V}^{(i)}$ for instance i (which is determined by each $\mathsf{p}^{(i)}$, and induces the corresponding distributions $\widehat{\pi}^{(i)}_h(a_{1:h}) \propto \pi_{\mathsf{ref},h}(a_{1:h}) \cdot \widehat{V}^{(i)}(a_{1:h})$). The ratios $V^\star(a_{1:h})/\widehat{V}^{(i)}(a_{1:h})$ which are needed to estimate $D_{\chi^2}\left(\pi^\star_h \,\|\, \widehat{\pi}^{(i)}_h\right)$ are prohibitively high in variance, so we instead estimate $D_{\mathsf{KL}}\left(\pi^\star_h \,\|\, \widehat{\pi}^{(i)}_h\right)$ to measure the divergence between $\pi^\star_h$ and $\widehat{\pi}^{(i)}_h$. We use Qwen3-0.6B as the language model M and set the horizon to $H = 64$ tokens. For each choice of $i \in [k]$ we run SMC with $N = 32$ particles given the PRM $\widehat{V}^{(i)}$, and sample from its output distribution. In Figure 1a, we plot the values of $D_{\mathsf{KL}}\left(\pi^\star_h \,\|\, \widehat{\pi}^{(i)}_h\right)$ against the sampling error of SMC, measured via $\mathcal{P}$-logprob discrepancies (estimated by running SMC for 200 trials). We observe a clear correlation between the two quantities, thus validating our theoretical findings that PRM accuracy predicts SMC's sampling error. In Appendix H.5 we repeat the experiment for additional choices of the various prompts and observe similar results.

We next evaluate the influence of action-level coverage on SMC performance. We consider the same setup as above, except for each instance $i \in [k]$ we take the target distribution, denoted $(\pi^\star)^{(i)}$, to be the same as the distribution $\widehat{\pi}^{(i)}$ induced by $\widehat{V}^{(i)}$, and fix $\mathsf{p}_{\mathsf{ref}}$ (and thus π_{ref}) across all instances. Doing so allows us to control for the divergence between $(\pi^\star)^{(i)}$ and $\widehat{\pi}^{(i)}$ while varying action-level coverage. In Figure 1b we observe that action-level coverage correlates with the sampling error of SMC. Here, since the worst-case quantity C_{act} is intractable to estimate, we measure action-level coverage via the average-case proxy $\frac{1}{H} D_{\mathsf{KL}}(\pi^\star \,\|\, \pi_{\mathsf{ref}})$, which is the average one-step log density ratio between $\pi^\star$ and π_{ref} (Appendix H.3). We refer the reader to Appendix H for further details.

4.3 SMC for Math Reasoning Tasks

Having validated that our criteria predict SMC's *sampling error* in a controlled setting, we now ask whether they also predict its *downstream accuracy* on math reasoning tasks (AIME and Math500), where each particle represents a complete chain of thought leading to an answer. In this setting, the base distribution π_{ref} corresponds to the output distribution of a language model prompted to solve a math problem, and the terminal reward function $r^\star(a_{1:H}) \in \{0, 1\}$ is the indicator that the answer $a_{1:H}$ is correct. Thus, $\pi^\star$ is simply the base model distribution *conditioned on producing a correct answer*, and $V^\star(a_{1:h})$ is the probability that π_{ref} produces a correct answer conditioned on $a_{1:h}$.

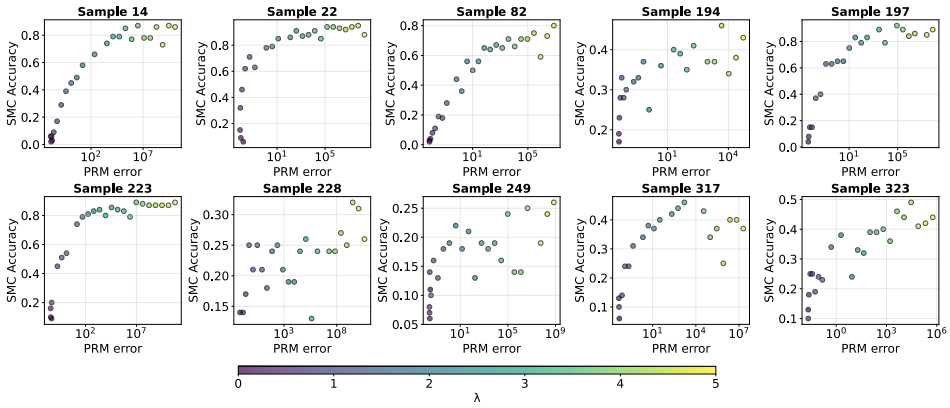


Figure 3: Influence of PRM error (measured by $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$) on SMC accuracy for 10 Math500 problems. Different colors correspond to different values of the inverse temperature parameter λ parametrizing $\widehat{V}^{(\lambda)}$. The plots show SMC accuracy with respect to a *random* particle (not the best one) selected at the end of SMC.

SMC improves performance on math. Many papers [6, 42] have reported that SMC typically improves the performance over Best-of-N (as well as other baselines). However, such results are typically reported *on-average* over all problems in the dataset, which fails to account for the heterogeneity between individual problems. For instance, it was not clear if SMC improves performance over Best-of-N on *all* problems, or if SMC can actually hurt performance (compared to Best-of-N) on a non-negligible fraction of the dataset. Certainly such a distinction is salient for any theoretical explanation for SMC's performance.

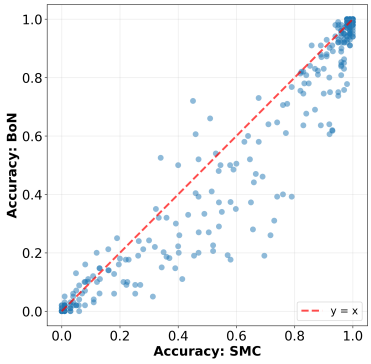


Figure 2: Performance of SMC (with N particles) vs. Best-of-N on Math500; here we take $N = 32$. Each point is a different problem; thus, SMC with N particles improves performance over Best-of-N on most problems. See Section 4.3.

In Figure 2, we estimate the performance of SMC and Best-of-N for each problem in Math500 (using 100 trials per problem), and plot the accuracy of SMC against that of Best-of-N. While there is variability in the relative performance of the two algorithms, we see that for the vast majority of problems, SMC's accuracy is at least as large as that of Best-of-N. In Figure 6, we observe a similar conclusion for AIME24 and AIME25. We note that SMC with N particles and Best-of-N sample the same number of actions (namely, NH) from $\pi_{\texttt{ref}}$, though SMC additionally queries the PRM on partial generations.

SMC and PRM accuracy. Having established that SMC leads to a consistent increase in accuracy across most Math500 and AIME problems, we next turn to the question of what factors influence its performance *on a given problem*. In light of the fact that significant effort has been spent designing better PRMs $\widehat{V}$ [32, 51, 52, 61], a natural follow-up question from our theory is whether the divergence between $\pi_h^\star$ and $\widehat{\pi}_h$ (as measured by, e.g., the χ^2-divergence $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$) predicts the performance of SMC.

To address this question, we consider a range of PRMs $\widehat{V}^{(\lambda)}$ by varying the inverse temperature parameter λ used by the PRM to produce its final probabilities. In Figure 3, we select 10 Math500 problems, and for each value of λ, we plot an empirical estimate of $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ against the accuracy of SMC given $\widehat{V}^{(\lambda)}$. Perhaps surprisingly, we observe that larger χ^2-divergence often leads to *higher* accuracy, the opposite of what our theory predicts. We suspect that larger values of λ (lower temperature) lead to better performance since they more effectively weed out incorrect partial solutions; however, they will also potentially lead $\widehat{\pi}_h$ to miss some modes of $\pi_h^\star$, which causes the χ^2-divergence to blow up. This suggests that perspectives beyond distributional sampling may be necessary to fully explain SMC's effectiveness on reasoning tasks.

"Sharpening": using the model's probabilities as a PRM. In some settings, a PRM may not be available, or may be too expensive to train. Therefore, it is of interest to explore alternatives to the use of bespoke PRMs. Accordingly, we consider the setting of *sharpening* $\pi_{\texttt{ref}}$, namely that where $\widehat{V}(a_{1:h}) := \pi_{\texttt{ref}}(a_{1:h})^{\lambda}$, for some $\lambda > 0$. Given that $V^{\star}(a_{1:H}) = \widehat{V}(a_{1:H}) = \pi_{\texttt{ref}}(a_{1:H})^{\lambda}$, it follows that $\pi^{\star}_{H}(a_{1:H}) \propto \pi_{\texttt{ref}}(a_{1:H})^{\lambda+1}$, so that $\pi^{\star}_{H}$ is a "sharpened" version of $\pi_{\texttt{ref}}$: responses $a_{1:H}$ which are more likely under $\pi_{\texttt{ref}}$ are further amplified under $\pi^{\star}_{H}$. Previous work has investigated the use of alternative techniques, such as Best-of-N [26, 27] and the Metropolis-Hastings algorithm [28] to sample from sharpened distributions. In Figure 4, we consider an experimental setup along these lines (see Appendix H.8 for full details); we observe that using SMC, where $\pi_{\texttt{ref}}$ is given by (a lower-temperature variant of) Qwen2.5-7B-Instruct, significantly improves the performance on Math500 problems over baseline. Further, using Best-of-N in this setup (i.e., drawing N responses $a_{1:H} \sim \pi_{\texttt{ref}}$ and choosing the one with the maximum value of $\pi_{\texttt{ref}}(a_{1:H})$) *degrades* performance, an "overoptimization" phenomenon which has been found in prior works as well [17, 27].

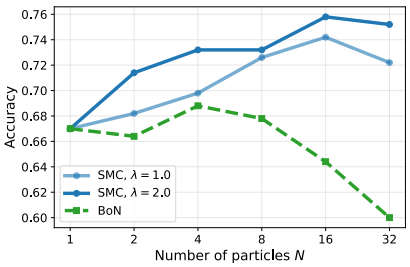


Figure 4: Accuracy of SMC and Best-of-N on Math500 in the sharpening setting, with $\pi_{\texttt{ref}}$ given by Qwen2.5-7B-Instruct.

Effective sample size. Our guarantees control the error of the output distribution of SMC averaged over the randomness of the algorithm, and do not by themselves rule out *particle degeneracy* within a single run, i.e., the effective sample size $(\sum_i w_h^i)^2 / \sum_i (w_h^i)^2$ of the weights in Algorithm 1 collapsing to 1. In our experiments, we observed that the effective sample size generally does not collapse to 1, which is consistent with the empirical effectiveness of SMC.

5 Conclusion

We developed a non-asymptotic analysis of multi-particle methods for steering language models with a PRM. Two natural criteria, action-level coverage and the χ^2-accuracy of the PRM, suffice for the convergence of SMC, and similar guarantees hold for a simple modification, SMC-RS. Empirically, our criteria predict the sampling error of SMC in a controlled setting, but not its accuracy on math reasoning benchmarks, suggesting that explaining the success of SMC on reasoning tasks may require perspectives beyond distributional sampling. We discuss further open directions in Appendix A.

Acknowledgments and Disclosure of Funding

This work received no third-party funding or support beyond the authors' listed affiliations. The authors declare no competing interests.

References

[1] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002. doi: 10.1109/78.978374.

[2] Edoardo Botta, Yuchen Li, Aashay Mehta, Jordan T. Ash, Cyril Zhang, and Andrej Risteski. On the query complexity of verifier-assisted language generation. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=9oIjvaDhoN.

[3] Bradley Brown, Jordan Juravsky, Ryan Ehrlich, Ronald Clark, Quoc V Le, Christopher Ré, and Azalia Mirhoseini. Large language monkeys: Scaling inference compute with repeated sampling. *arXiv preprint arXiv:2407.21787*, 2024.

[4] Olivier Cappé, Simon J. Godsill, and Eric Moulines. An overview of existing methods and recent advances in sequential monte carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007. doi: 10.1109/JPROC.2007.893250.

[5] F Cérou, P Del Moral, and A Guyader. A nonasymptotic theorem for unnormalized feynman–kac particle models. In *Annales de l'IHP Probabilités et statistiques*, volume 47, pages 629–649. Gauthier-Villars, 2011.

[6] Giannis Chatziveroglou. A*-decoding: Token-efficient inference scaling, 2025. URL https://arxiv.org/abs/2505.13672.

[7] Fan Chen, Audrey Huang, Noah Golowich, Sadhika Malladi, Adam Block, Jordan T Ash, Akshay Krishnamurthy, and Dylan J Foster. The coverage principle: How pre-training enables post-training. In *International Conference on Learning Representations*, 2026. Oral presentation.

[8] Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer, Cham, Switzerland, 2020. doi: 10.1007/978-3-030-47845-2.

[9] Pierre Del Moral. Feynman-kac formulae. In *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, pages 47–93. Springer, 2004.

[10] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006. doi: 10.1111/j.1467-9868.2006.00553.x.

[11] Randal Douc and Olivier Cappé. Comparison of resampling schemes for particle filtering. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 64–69, 2005. doi: 10.1109/ISPA.2005.195385.

[12] Arnaud Doucet and Adam M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. Technical report, Department of Statistics, University of Warwick, 2009.

[13] Arnaud Doucet, Simon J. Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[14] Shengyu Feng, Xiang Kong, Shuang Ma, Aonan Zhang, Dong Yin, Chong Wang, Ruoming Pang, and Yiming Yang. Step-by-step reasoning for math problems via twisted sequential monte carlo. In *International Conference on Learning Representations*, 2025.

[15] Dylan J. Foster, Zakaria Mhammedi, and Dhruv Rohatgi. Is a good foundation necessary for efficient reinforcement learning? the computational role of the base model in exploration. *arXiv preprint arXiv:2503.07453*, 2025.

[16] Yichao Fu, Xuewei Wang, Yuandong Tian, and Jiawei Zhao. Deep think with confidence. In *NeurIPS 2025 Workshop on Efficient Reasoning*, 2025.

[17] Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 10835–10866. PMLR, 23–29 Jul 2023.

[18] Mathieu Gerber, Nicolas Chopin, and Nick Whiteley. Negative association, ordering and convergence of resampling methods. *The Annals of Statistics*, 47(4):2236–2260, 2019. doi: 10.1214/18-AOS1746.

[19] Jonathan Geuter, Youssef Mroueh, and David Alvarez-Melis. Guided speculative inference for efficient test-time alignment of LLMs. *arXiv preprint arXiv:2506.04118*, 2025.

[20] Noah Golowich, Allen Liu, and Abhishek Shetty. Sequences of logits reveal the low rank structure of language models, 2025. URL https://arxiv.org/abs/2510.24966.

[21] Neil J. Gordon. Non-linear/non-gaussian filtering and the bootstrap filter. In *IEE Colloquium on Non-linear Filters*. IET, 1994.

[22] Gabriel Grand, Joshua B. Tenenbaum, Vikash K. Mansinghka, Alexander K. Lew, and Jacob Andreas. Self-steering language models. In *Conference on Language Modeling (COLM)*, 2025. arXiv:2504.07081.

[23] Peter Guarniero, Adam M. Johansen, and Anthony Lee. The iterated auxiliary particle filter. *Journal of the American Statistical Association*, 112(520):1636–1647, 2017.

[24] Martin Hairer and Jonathan Weare. Improved diffusion monte carlo. *Communications on Pure and Applied Mathematics*, 67(12):1995–2021, 2014.

[25] Jeremy Heng, Adrian Bishop, George Deligiannidis, and Arnaud Doucet. Controlled sequential monte carlo. *The Annals of Statistics*, 48(5):2904–2929, 2020. doi: 10.1214/19-AOS1914.

[26] Audrey Huang, Adam Block, Dylan J. Foster, Dhruv Rohatgi, Cyril Zhang, Max Simchowitz, Jordan T. Ash, and Akshay Krishnamurthy. Self-improvement in language models: The sharpening mechanism. *arXiv preprint arXiv:2412.01951*, 2024.

[27] Audrey Huang, Adam Block, Qinghua Liu, Nan Jiang, Akshay Krishnamurthy, and Dylan J. Foster. Is best-of-n the best of them? coverage, scaling, and optimality in inference-time alignment. *arXiv preprint arXiv:2503.21878*, 2025.

[28] Aayush Karan and Yilun Du. Reasoning with sampling: Your base model is smarter than you think, 2025. URL https://arxiv.org/abs/2510.14901.

[29] Dieterich Lawson, Allan Raventós, Andrew Warrington, and Scott Linderman. Sixo: Smoothing inference with twisted objectives. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

[30] Holden Lee and Matheau Santana-Gijzen. Convergence bounds for sequential monte carlo on multimodal distributions using soft decomposition. *arXiv preprint arXiv:2405.19553*, 2024.

[31] Alexander K. Lew, Zhi-Xuan Tan, Gabriel Grand, and Vikash K. Mansinghka. Sequential monte carlo steering of large language models using probabilistic programs. *arXiv preprint arXiv:2306.03081*, 2023.

[32] Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. *arXiv preprint arXiv:2305.20050*, 2023.

[33] Benjamin Lipkin, Benjamin LeBrun, Jacob Hoover Vigly, João Loula, David R. MacIver, Li Du, Jason Eisner, Ryan Cotterell, Vikash K. Mansinghka, Timothy J. O'Donnell, Alexander K. Lew, and Tim Vieira. Fast controlled generation from language models with adaptive weighted rejection sampling. *arXiv preprint arXiv:2504.05410*, 2025.

[34] João Loula, Benjamin LeBrun, Li Du, Ben Lipkin, Clemente Pasti, Gabriel Grand, Tianyu Liu, Yahya Emara, Marjorie Freedman, Jason Eisner, Ryan Cotterell, Vikash K. Mansinghka, Alexander K. Lew, Tim Vieira, and Timothy J. O'Donnell. Syntactic and semantic control of large language models via sequential monte carlo. In *International Conference on Learning Representations*, 2025.

[35] Lovish Madaan, Aniket Didolkar, Suchin Gururangan, John Quan, Ruan Silva, Ruslan Salakhutdinov, Manzil Zaheer, Sanjeev Arora, and Anirudh Goyal. Rethinking thinking tokens: Llms as improvement operators. *arXiv preprint arXiv:2510.01123*, 2025.

[36] Joseph Marion and Scott C Schmidler. Finite sample l_2 bounds for sequential monte carlo and adaptive path selection. *arXiv preprint arXiv:1807.01346*, 2018.

[37] Michael Montemerlo, Sebastian Thrun, Daphne Koller, and Ben Wegbreit. Fastslam: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2002.

[38] Sidharth Mudgal, Jong Lee, Harish Ganapathy, YaGuang Li, Tao Wang, Yanping Huang, Zhifeng Chen, Heng-Tze Cheng, Michael Collins, Trevor Strohman, Jilin Chen, Alex Beutel, and Ahmad Beirami. Controlled decoding from language models. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=bVIcZb7Qa0.

[39] C. A. Naesseth, F. Lindsten, and T. B. Schön. High-dimensional filtering using nested sequential monte carlo. *IEEE Transactions on Signal Processing*, 67(16):4177–4188, 2019. doi: 10.1109/TSP.2019.2926035.

[40] Michael K. Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. Technical Report W13, Nuffield College, University of Oxford, 1997.

[41] Isha Puri, Shivchander Sudalairaj, Guangxuan Xu, Kai Xu, and Akash Srivastava. A probabilistic inference approach to inference-time scaling of LLMs using particle-based monte carlo methods. *arXiv preprint arXiv:2502.01618*, 2025.

[42] Isha Puri, Shivchander Sudalairaj, Guangxuan Xu, Kai Xu, and Akash Srivastava. A probabilistic inference approach to inference-time scaling of llms using particle-based monte carlo methods. *arXiv preprint arXiv:2502.01618*, 2025.

[43] Dhruv Rohatgi, Abhishek Shetty, Donya Saless, Yuchen Li, Ankur Moitra, Andrej Risteski, and Dylan J Foster. Taming imperfect process verifiers: A sampling perspective on backtracking. *arXiv preprint arXiv:2510.03149*, 2025.

[44] Mathieu Rousset. *Méthodes de Monte Carlo en temps continu pour la physique numérique*. Ph.d. thesis, Université Paul Sabatier (Toulouse III), Toulouse, France, 2006.

[45] Nikolaus Schweizer. Non-asymptotic error bounds for sequential mcmc methods in multimodal settings. *arXiv preprint arXiv:1205.6733*, 2012.

[46] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and computation*, 82(1):93–133, 1989.

[47] Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. *arXiv preprint arXiv:2501.06848*, 2025.

[48] Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alán Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *arXiv preprint arXiv:2503.02819*, 2025.

[49] Andreas Stephan, Dawei Zhu, Matthias Assenmacher, Xiaoyu Shen, and Benjamin Roth. From calculation to adjudication: Examining LLM judges on mathematical reasoning tasks. *arXiv preprint arXiv:2409.04168*, 2024.

[50] Kaiwen Wang, Jin Peng Zhou, Jonathan Chang, Zhaolin Gao, Nathan Kallus, Kianté Brantley, and Wen Sun. Value-guided search for efficient chain-of-thought reasoning. *arXiv preprint arXiv:2505.17373*, 2025.

[51] Kaiwen Wang, Jin Peng Zhou, Jonathan Chang, Zhaolin Gao, Nathan Kallus, Kiante Brantley, and Wen Sun. Value-guided search for efficient chain-of-thought reasoning. In *Advances in Neural Information Processing Systems*, 2025. arXiv:2505.17373.

[52] Peiyi Wang, Lei Li, Zhihong Shao, R. X. Xu, Damai Dai, Yifei Li, Deli Chen, Y. Wu, and Zhifang Sui. Math-shepherd: Verify and reinforce LLMs step-by-step without human annotations. *arXiv preprint arXiv:2312.08935*, 2024.

[53] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=1PL1NIMMrw.

[54] Nick Whiteley, Nikolas Kantas, and Ajay Jasra. Linear variance bounds for particle approximations of time-homogeneous feynman–kac formulae. *Stochastic Processes and their Applications*, 122(4):1840–1865, 2012.

[55] Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pages 31372–31403, 2023.

[56] Xixi Wu, Kuan Li, Yida Zhao, Liwen Zhang, Litu Ou, Huifeng Yin, Zhongwang Zhang, Xinmiao Yu, Dingchu Zhang, Yong Jiang, et al. Resum: Unlocking long-horizon search intelligence via context summarization. *arXiv preprint arXiv:2509.13313*, 2025.

[57] Wei Xiong, Hanze Dong, Chenlu Ye, Ziqi Wang, Han Zhong, Heng Ji, Nan Jiang, and Tong Zhang. Iterative preference learning from human feedback: Bridging theory and practice for RLHF under KL-constraint. *arXiv preprint arXiv:2312.11456*, 2024.

[58] Kevin Yang and Dan Klein. Fudge: Controlled text generation with future discriminators. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3511–3535, 2021.

[59] Ling Yang, Zhaochen Yu, Tianjun Zhang, Shiyi Cao, Minkai Xu, Wentao Zhang, Joseph E Gonzalez, and Bin Cui. Buffer of thoughts: Thought-augmented reasoning with large language models. *Advances in Neural Information Processing Systems*, 37:113519–113544, 2024.

[60] Yue Yu, Qiwei Di, Quanquan Gu, and Dongruo Zhou. On the limits of test-time compute: Sequential reward filtering for better inference. *arXiv preprint arXiv:2512.04558*, 2025.

[61] Zhenru Zhang, Chujie Zheng, Yangzhen Wu, Beichen Zhang, Runji Lin, Bowen Yu, Dayiheng Liu, Jingren Zhou, and Junyang Lin. The lessons of developing process reward models in mathematical reasoning. *arXiv preprint arXiv:2501.07301*, 2025.

[62] Stephen Zhao, Rob Brekelmans, Alireza Makhzani, and Roger Grosse. Probabilistic inference in language models via twisted sequential monte carlo. *arXiv preprint arXiv:2404.17546*, 2024.

[63] Wenting Zhao, Pranjal Aggarwal, Swarnadeep Saha, Asli Celikyilmaz, Jason Weston, and Ilia Kulikov. The majority is not always right: Rl training for solution aggregation. *arXiv preprint arXiv:2509.06870*, 2025.

[64] Youheng Zhu and Yiping Lu. On the power of (approximate) reward models for inference-time scaling. *arXiv preprint arXiv:2602.01381*, 2026.

Contents of Appendix

A Discussion **16**

B Comparison with Prior Work **16**

C Technical Tools **17**

D Proofs for SMC **17**
D.1 Useful Lemmas for SMC 17
D.2 Proof of Theorem 3.2 19
D.3 Insufficiency of Existing Analysis 20
D.4 Proof of Theorem 3.5 21
D.5 Proof of Theorem 3.6 22
D.6 Lower Bound for SMC with Exact Value Function 23

E Proofs for SMC-RS **25**
E.1 Proof of Theorem 3.8 26
E.2 Analysis of SMC-RS with Heavy-Tailed Errors 27
E.3 Faster Convergence for SMC-RS 27

F A Unified Perspective on Backtracking and Multi-Particle Methods **29**
F.1 Background on VGB 29
F.2 Definition of SMC-IND 30
F.3 Coupling VGB with SMC-IND 30

G Proof of Theorem 3.9 **31**
G.1 Myopic Multi-Particle Methods 31
G.2 Specialized Setting and Formal Theorem Statement 31
G.3 Proof of Theorem G.5 32

H Additional Experimental Details **38**
H.1 Additional Evaluation Approaches 38
H.2 PRM Accuracy Predicts SMC Performance 38
H.3 Action-Level Coverage Predicts SMC Performance 39
H.4 Scaling with the Number of Particles 40
H.5 Details for Appendix H.2 40
H.6 Details for Appendix H.3 41
H.7 SMC vs. Best-of-N with Qwen Math PRM 41
H.8 Details for SMC on Math with Sharpening setup 41

A Discussion

In this paper, we have established theoretical guarantees for multi-particle methods under various assumptions relating the reference distribution $\pi_{\mathtt{ref}}$ to the target distribution $\pi^\star$ and the PRM $\widehat{V}$ (via the induced distributions $\widehat{\pi}$). We highlight below some intriguing technical directions for future work:

- *Closing polynomial gaps.* There are various polynomial gaps between our upper and lower bounds, including the dependence on horizon H, as discussed in Section 1.2, as well as the decay of the total variation distance as a function of N: whereas our upper bounds of Theorems 3.2 and 3.8 decay as $1/\sqrt{N}$, the lower bound of Proposition D.8 decays as only $1/N$.
- *Weaker notions of distributional closeness.* Can we develop metrics for the performance of SMC (and other inference-time sampling algorithms) which are weaker than total variation distance but still meaningfully capture "usefulness" of outputs? For instance, for math problems, typically one only desires a correct answer, and the particular phrasing of the solution may not matter too much.
- *Beyond χ^2-divergence.* Can we develop provable guarantees for multi-particle methods or other types of inference-time sampling procedures which rely on weaker notions of discrepancy between $\pi^\star_h$ and $\widehat{\pi}_h$? The values of $D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h)$ seen in, e.g., the experiments in Appendices H.2 to H.4 are extremely large (in particular, much larger than the values of N seen in our experiments), yet multi-particle methods still are able to effectively make use of $\widehat{V}$.
- *Additional variants of SMC & multi-particle methods.* The version of SMC we consider in Algorithm 1 uses multinomial resampling, which is known to be suboptimal in terms of variance reduction. Can we get better guarantees for SMC with more sophisticated resampling schemes such as stratified resampling [11, 18]? Another direction is to investigate the benefit of *particle smoothing* approaches [4], which incorporate a "backward pass" over the Markov chain and may be related to some of the empirical benefits of backtracking observed in prior work [43].

B Comparison with Prior Work

In this section we discuss how our theoretical guarantees compare with prior work. We specify to the autoregressive generation setting, as the more general setting has not been theoretically studied in the past.

Action-level rejection sampling / sequential importance sampling [58]. This algorithm samples $a_{1:H}$ autoregressively via the conditional distributions

$$\widehat{\pi}(a_{h+1} \mid a_{1:h}) \propto \pi_{\mathtt{ref}}(a_{1:h+1} \mid a_{1:h})\widehat{V}(a_{1:h+1}).$$

It can be implemented in expected time $O(H\widehat{C}_{\mathsf{act}})$ by rejection sampling / importance sampling [43], and it exactly samples from $\pi^\star$ if $\widehat{V} = V^\star$. However, even under assumption Assumption 3.3 with $\mathsf{C}_\infty = 1 + \epsilon$, it can incur $\Omega(\epsilon\sqrt{H})$ sampling error [43], which is vacuous if e.g. $\epsilon = \Theta(1)$.

VGB [43]. This algorithm is defined by a random walk on the *tree of autoregressive generations*, which has state space $\mathcal{A}^0 \sqcup \cdots \sqcup \mathcal{A}^H$ and in which the parent of $a_{1:h}$ is $a_{1:h-1}$. The first main result of Rohatgi et al. [43] is that the random walk samples from $\pi^\star$, up to total variation error δ, in time roughly $O(H^2 C_{\mathsf{act}} \mathsf{C}_\infty^4 \log(1/\delta))$. This guarantee is essentially matched by the guarantee of Theorem 3.6 for SMC, except the language model evaluations in SMC can be parallelized into H rounds, whereas the language model evaluations in VGB are inherently sequential.[6]

The second main result of Rohatgi et al. [43] shows that under the following average-case closeness condition between $\widehat{V}$ and $V^\star$, as well as bounded action-level coverage, VGB samples from a distribution that *approximately covers* $\pi^\star$.

[6] The results of Appendix F suggest a potential approach to parallelizing VGB via multi-particle methods, but it is unclear in what generality such a simulation would be efficient.

Assumption B.1 (Average-case assumption of Rohatgi et al. [43])**.**

$$\max\left\{\mathbb{E}^{\pi^\star}\left[\frac{\widehat{V}(a_{1:h})}{V^\star(a_{1:h})}\right], \mathbb{E}^{\pi^\star}\left[\frac{V^\star(a_{1:h})}{\widehat{V}(a_{1:h})}\right]\right\} \le C.$$

We make the following observation:

Fact B.2. *For each $0 \le h \le H$, it holds that*

$$D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) = \mathbb{E}^{\pi^\star}\left[\frac{\widehat{V}(a_{1:h})}{V^\star(a_{1:h})}\right] \cdot \mathbb{E}^{\pi^\star}\left[\frac{V^\star(a_{1:h})}{\widehat{V}(a_{1:h})}\right] - 1.$$

Thus, under Assumption B.1 as well as bounded action-level coverage, Theorem 3.2 shows that SMC with $N \ge (C+1)^2 H^2/\epsilon^2$ particles samples from a distribution that is ϵ-close in total variation distance to $\pi^\star$. Note that closeness in total variation distance is a stronger guarantee than approximate coverage.

C Technical Tools

Proposition C.1. *For any $R \ge 0$, independent random variables $X_1, \cdots, X_n$, it holds that*

$$\mathbb{E}\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \le \sqrt{nR \max_i \mathbb{E}|X_i|} + 2n \max_i \mathbb{E}[|X_i| \mathbb{I}\{|X_i| \ge R\}].$$

Proof. We define $Y_i = \min\{\max\{X_i, -R\}, R\}$ and $Z_i = X_i - Y_i$. Then, we can bound

$$\mathbb{E}\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \le \mathbb{E}\left|\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])\right| + \sum_{i=1}^n \mathbb{E}|Z_i - \mathbb{E}[Z_i]|.$$

Note that $\mathbb{E}|Z_i - \mathbb{E}[Z_i]| \le 2\,\mathbb{E}|Z_i| \le 2\,\mathbb{E}[|X_i| \mathbb{I}\{|X_i| \ge R\}]$. Further,

$$\begin{aligned}\left(\mathbb{E}\left|\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])\right|\right)^2 &\le \mathbb{E}\left(\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])\right)^2 \\ &= \sum_{i=1}^n \mathbb{E}(Y_i - \mathbb{E}[Y_i])^2 \le \sum_{i=1}^n \mathbb{E}\, Y_i^2 \le R \sum_{i=1}^n \mathbb{E}|X_i|,\end{aligned}$$

where we use $|Y_i| \le \min\{|X_i|, R\}$. The proof is then completed by combining the inequalities above. □

D Proofs for SMC

Notation. We define $\mathcal{H}_h = \{(\tilde{x}_\ell^i, x_\ell^i)\}_{\ell \in [h], i \in [N]}$ to be the full history of the execution of SMC (Algorithm 1) up to the end of iteration h. We write $\mathbb{E}_h[\cdot]$ for the conditional distribution with respect to $\mathcal{H}_h$. We write $Z := \mathbb{E}_{x_{1:H} \sim \pi_{\mathsf{ref}}}[r^\star(x_H)]$ and observe that $Z = \mathbb{E}_{x \sim \pi_h}[V^\star(x)]$ for all h.

D.1 Useful Lemmas for SMC

We first prove the following lemmas.

Lemma D.1. *For any step h and any function $g : \mathcal{X}_h \to \mathbb{R}$, we have that*

$$\mathbb{E}_{h-1}\left[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g]\right] = \widehat{W}_{h-1} \cdot \mathbb{E}_{x_{h-1} \sim \widehat{\nu}_{h-1}}\left[\frac{\mathbb{E}_{x_h \sim \pi_{\mathsf{ref}}(\cdot | x_{h-1})}\left[\widehat{V}(x_h) \cdot g(x_h)\right]}{\widehat{V}(x_{h-1})}\right]. \tag{1}$$

The next lemma follows from applying Lemma D.1 inductively.

Lemma D.2. *For any $g_h : \mathcal{X}_h \to \mathbb{R}$, it holds that*

$$\mathbb{E}\Big[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g_h]\Big] = \mathbb{E}_{x_h \sim \pi_h}\Big[\widehat{V}(x_h) g_h(x_h)\Big]. \tag{2}$$

In particular, $\mathbb{E}[\widehat{W}_h] = \widehat{Z}_h = \mathbb{E}[\widehat{V}(x_h)]$ for $h \geq 1$ and

$$\mathbb{E}\Big[\widehat{W}_H \widehat{\nu}_H(x)\Big] = Z \cdot \pi^\star_H(x), \qquad \forall x \in \mathcal{X}_H. \tag{3}$$

Finally, we upper bound the TV error of Algorithm 1 by the deviation of $\widehat{W}_H$.

Lemma D.3. *For Algorithm 1, it holds that*

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq \frac{1}{2Z}\,\mathbb{E}|\widehat{W}_H - Z|.$$

Proof of Lemma D.1 and Lemma D.2. To prove Eq. (1), observe that

$$\begin{aligned}
&\mathbb{E}_{h-1}\Big[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g]\Big] \\
&= \widehat{W}_{h-1} \cdot \mathbb{E}_{h-1}\left[\frac{1}{N}\sum_{i=1}^N w_h^i g(x_h^i)\right] \\
&= \widehat{W}_{h-1} \cdot \mathbb{E}_{h-1}\left[\frac{1}{N}\sum_{i=1}^N \frac{\widehat{V}(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)} g(x_h^i)\right].
\end{aligned}$$

Since $\big\{(\tilde{x}_{h-1}^i, x_h^i)\big\}$ are i.i.d. given $\mathcal{H}_{h-1}$, for each $i \in [N]$, we have

$$\mathbb{E}_{h-1}\left[\frac{\widehat{V}(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)} g(x_h^i)\right] = \mathbb{E}_{\tilde{x}_{h-1} \sim \hat{\nu}_{h-1}} \mathbb{E}_{x_h \sim \pi_{\mathsf{ref}}(\cdot|\tilde{x}_{h-1})}\left[\frac{\widehat{V}(x_h)}{\widehat{V}(\tilde{x}_{h-1})} g(x_h)\right],$$

which establishes Eq. (1). Next, we argue that Eq. (2) holds inductively (noting that the base case is trivial since $\widehat{W}_0 = \widehat{V}(\perp)$ and $\widehat{\nu}_0 = \delta_\perp$), using Eq. (1):

$$\begin{aligned}
\mathbb{E}\Big[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g_h]\Big] &= \mathbb{E}\left[\widehat{W}_{h-1} \cdot \mathbb{E}_{x_{h-1} \sim \widehat{\nu}_{h-1}}\left[\frac{\mathbb{E}_{x_h \sim \pi_{\mathsf{ref}}(\cdot|x_{h-1})}\Big[\widehat{V}(x_h) \cdot g_h(x_h)\Big]}{\widehat{V}(x_{h-1})}\right]\right] \\
&= \mathbb{E}\left[\widehat{V}(x_{h-1}) \cdot \frac{\mathbb{E}_{x_h \sim \pi_{\mathsf{ref}}(\cdot|x_{h-1})}[\widehat{V}(x_h) g_h(x_h)]}{\widehat{V}(x_{h-1})}\right] \\
&= \mathbb{E}[\widehat{V}(x_h) g_h(x_h)],
\end{aligned}$$

as desired. We get that $\mathbb{E}[\widehat{W}_h] = \mathbb{E}[\widehat{V}(x_h)]$ by using $g_h \equiv 1$. Finally, Eq. (3) follows by setting $g_H(x) = \mathbb{I}\{x_H = x\}$ in Eq. (2) and using that $\widehat{V}(x_H) = V^\star(x_H)$ for $x_H \in \mathcal{X}_H$. □

Proof of Lemma D.3. By Eq. (3), we can write

$$\begin{aligned}
2Z D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) &= \sum_{x \in \mathcal{X}_H} |Z\,\mathbb{E}[\widehat{\nu}_H(x)] - Z\pi^\star_H(x)| = \sum_{x \in \mathcal{X}_H} \Big|Z\,\mathbb{E}[\widehat{\nu}_H(x)] - \mathbb{E}[\widehat{W}_H \widehat{\nu}_H(x)]\Big| \\
&\leq \sum_{x \in \mathcal{X}_H} \mathbb{E}\Big[\Big|\widehat{W}_H - Z\Big|\widehat{\nu}_H(x)\Big] = \mathbb{E}\left[\Big|\widehat{W}_H - Z\Big| \cdot \sum_{x \in \mathcal{X}_H} \widehat{\nu}_H(x)\right] = \mathbb{E}\Big|\widehat{W}_H - Z\Big|.
\end{aligned}$$

□

D.2 Proof of Theorem 3.2

We prove a slightly stronger upper bound.

Proposition D.4. *SMC (Algorithm 1) with N particles achieves*

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \le \frac{1}{2Z} \cdot \frac{1}{\sqrt{N}} \sum_{h=1}^{H} \sqrt{\mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\, \mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_{h-1})}\right]}.$$

Theorem 3.2 then follows from the fact that $V^\star(x_h) \le C_{\mathsf{act}} V^\star(x_{h-1})$ almost surely under $x_h \sim \pi_{\mathsf{ref}}(\cdot \mid x_{h-1})$ and

$$\mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\, \mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_{h-1})^2}{\widehat{V}(x_{h-1})}\right] = Z^2 \cdot \left(1 + D_{\chi^2}\left(\pi^\star_{h-1} \,\|\, \widehat{\pi}_{h-1}\right)\right).$$

□

Proof of Proposition D.4. We define $\rho(x) = \frac{V^\star(x)}{\widehat{V}(x)}$ and

$$M_h = \widehat{W}_h \cdot \mathbb{E}_{x_h \sim \widehat{\nu}_h}[\rho(x_h)].$$

Then Eq. (1) with $g(x) = \rho(x)$ implies (M_h) is a martingale with $M_0 = Z$ and $M_H = \widehat{W}_H$. We can then decompose

$$\mathbb{E}|\widehat{W}_H - Z| \le \sum_{h=1}^{H} \mathbb{E}|M_h - M_{h-1}|.$$

Next, we note that

$$\begin{aligned} M_h - M_{h-1} &= \widehat{W}_{h-1}\left(\frac{1}{N}\sum_{i=1}^{N} w^i_h \rho(x^i_h) - \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right) \\ &= \widehat{W}_{h-1}\left(\frac{1}{N}\sum_{i=1}^{N} \frac{V^\star(x^i_h)}{\widehat{V}(\tilde{x}^i_{h-1})} - \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right). \end{aligned}$$

Since the tuples $(\tilde{x}^i_{h-1}, x^i_h)$ are i.i.d. given $\mathcal{H}_{h-1}$, we can bound

$$\begin{aligned} \frac{1}{\widehat{W}^2_{h-1}} \mathbb{E}_{h-1}(M_h - M_{h-1})^2 &\le \frac{1}{N} \mathrm{Var}_{x_{h-1}\sim\widehat{\nu}_{h-1}, x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\right] \\ &\le \frac{1}{N}\, \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\frac{\mathbb{E}_{x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}[V^\star(x_h)^2]}{\widehat{V}(x_{h-1})^2}\right] =: \Delta_{h-1}. \end{aligned}$$

Then, by Lemma D.2, it holds that

$$\mathbb{E}[\widehat{W}_{h-1}]\, \mathbb{E}[\widehat{W}_{h-1}\Delta_{h-1}] = \frac{1}{N}\, \mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\, \mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_{h-1})}\right].$$

Therefore, we can bound

$$\begin{aligned} \mathbb{E}|\widehat{W}_H - Z| &\le \sum_{h=1}^{H} \mathbb{E}[\mathbb{E}_{h-1}|M_h - M_{h-1}|] \\ &\le \sum_{h=1}^{H} \mathbb{E}\left[\widehat{W}_{h-1}\sqrt{\Delta_{h-1}}\right] \le \sum_{h=1}^{H} \sqrt{\mathbb{E}[\widehat{W}_{h-1}] \cdot \mathbb{E}\left[\widehat{W}_{h-1}\Delta_{h-1}\right]} \\ &\le \frac{1}{\sqrt{N}} \sum_{h=1}^{H} \sqrt{\mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\, \mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_{h-1})}\right]}. \end{aligned}$$

Combining the inequalities above completes the proof. □

D.3 Insufficiency of Existing Analysis

Most non-asymptotic analyses of SMC assume conditions at least as strong as the L_∞ closeness in Assumption 3.3 [5, 9, 54]. Another line of work [30, 36] imposes assumptions on the transition kernel, which are therefore incomparable to those in this paper.

We note that these approaches might fail to capture SMC convergence when we assume only χ^2-closeness (Theorem 3.2). This is because many explicitly or implicitly control the variance $(\mathrm{Var}(\widehat{W}_h))_{h\in[H]}$ [5, 45, 54, etc.]. In contrast, our proof of Theorem 3.2 proceeds via the absolute deviation $\mathbb{E}|M_H - Z|$. The seemingly more natural attempt to study the variance $\mathbb{E}(M_H - Z)^2$ must confront the fact that $M_H = \widehat{W}_H = \prod_{h=1}^H \left(\frac{1}{N}\sum_{i=1}^N w_h^i\right)$ is a sequential product. As the following lemma suggests, $\mathrm{Var}(\widehat{W}_h)$ need not be controlled when the χ^2 divergences $(D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h))_{h\in[H]}$ are bounded.

Lemma D.5. *Suppose that $\widehat{V}(\perp) = 1$. For any $h \in [H]$, it holds that*

$$\mathrm{Var}(\widehat{W}_h) = \mathbb{E}(\widehat{W}_h - \widehat{Z}_h)^2 \geq \frac{\mathrm{Var}_{x_1\sim\pi_1}(\mathbb{E}[\widehat{V}(x_h) \mid x_1])}{N}.$$

In particular, for $H \geq 2$, there exists a problem instance $(V^\star, \widehat{V})$ such that $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) \leq 2$, and both Assumption 3.1 and Assumption 3.7 holds with $C_{\mathsf{act}} = 2$, but SMC with N particles incurs

$$\max_{h\in[H]} \mathrm{Var}(\widehat{W}_h) \geq \frac{2^{\lfloor H/2\rfloor} - 2}{N}.$$

Therefore, if convergence of SMC is established via upper bounds on $\mathrm{Var}(\widehat{W}_h)$, the rate can be vacuous unless we have $N \geq \exp(\Omega(H))$ particles. For example, Schweizer [45] study SMC through the following relation (cf. Lemma 2.1 therein) for $f : \mathcal{X}_H \to [0, 1]$:

$$\mathbb{E}\big(\mathbb{E}_{\widehat{\nu}_H}[f] - \mathbb{E}_{\pi_H^\star}[f]\big)^2 \leq 2\mathrm{Var}(\widehat{W}_H \cdot \mathbb{E}_{\widehat{\nu}_H}[f]) + 2\mathrm{Var}(\widehat{W}_H).$$

Consequently, Schweizer [45] establish upper bounds of the form

$$N \cdot \mathrm{Var}(\widehat{W}_H \cdot \mathbb{E}_{\widehat{\nu}_H}[f]) \leq C_0(f) + C_1(f) \cdot \epsilon_N,$$

where C_0, C_1 are two problem-dependent functionals, and ϵ_N is defined so that

$$\epsilon_N \geq \max_{h\in[H]} \mathrm{Var}(\widehat{W}_h).$$

This implies that the analysis in Schweizer [45] cannot recover the upper bound of Theorem 3.2 (and Theorem 3.5).

Proof of Lemma D.5. For any $h \leq \ell$, we define $\widetilde{V}_\ell(x_h) = \mathbb{E}[\widehat{V}(x_\ell) \mid x_h]$ and $\rho_\ell(x_h) = \frac{\widetilde{V}_\ell(x_h)}{\widehat{V}(x_h)}$, and we define $M_{h,\ell} = \widehat{W}_h\, \mathbb{E}_{\widehat{\nu}_h}[\rho_\ell]$. Then it is straightforward to verify that $(M_{h,\ell})_{h\in[\ell]}$ is a martingale with $M_{0,\ell} = \widehat{Z}_\ell$ and $M_{\ell,\ell} = \widehat{W}_\ell$.

Therefore, we can lower bound

$$\begin{aligned}
\mathbb{E}(\widehat{W}_\ell - \widehat{Z}_\ell)^2 &\geq \mathbb{E}(M_{1,\ell} - \widehat{Z}_\ell)^2 \\
&= \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N \frac{\widetilde{V}_\ell(x_1^i)}{\widehat{V}(\tilde{x}_0^i)} - \mathbb{E}_{\widehat{\nu}_0}[\rho_\ell]\right)^2\right] \\
&= \frac{1}{N}\Big(\mathbb{E}_{x_1\sim\pi_1}[\widetilde{V}_\ell(x_1)^2] - \widehat{Z}_\ell^2\Big),
\end{aligned}$$

where we note that $\tilde{x}_0^i = \perp$ for $i \in [N]$.

Now, consider $n = \lfloor H/2\rfloor$. Consider the following simple example:

- For each $h \in [H]$, $\mathcal{X}_h = \{(h, 0), (h, 1)\}$, and $P((h + 1, i) \mid (h, i)) = 1$ for $i \in \{0, 1\}$.

- The initial distribution is given by $\pi_1((1,0)) = p = 1 - \pi_1((1,1))$ for $p = 2^{-n}$.
- The reward function is $r \equiv 1$, and hence $V^\star(x) = 1$ for all $x \in \mathcal{X}$.
- For each $h \in [H]$, we set $\widehat{V}((h,1)) = 1$ and $\widehat{V}((h,0)) = 2^h$ for $h \le n$, and $\widehat{V}((h,0)) = 2^{(2n-h)_+}$ for $h \ge n$.

Then, it is clear that both Assumption 3.1 and Assumption 3.7 hold with $C_{\mathsf{act}} = 2$, and $\widehat{\pi}_h$ is given by $\widehat{\pi}((h,0)) = \frac{p\widehat{V}((h,0))}{p\widehat{V}((h,0))+1-p}$, and hence $\widehat{\pi}((h,0)) \in [p, \frac{1}{2}]$. This immediately implies $D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h) \le 2$ for any $h \in [H]$.

However, we can then calculate $\widetilde{V}_n((1,0)) = 2^n$, and hence $\mathbb{E}[\widetilde{V}_n(x_1)^2] \ge p \cdot 2^{2n} = 2^n$. This gives the desired statement. □

D.4 Proof of Theorem 3.5

We prove the following general version of Theorem 3.5.

Theorem D.6 (SMC convergence under coverage (full version))**.** *SMC with N particles achieves the following for any $M \ge 1$, $\eta \ge 1$,*

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \le H\sqrt{\frac{M\eta}{N}} + \sum_{h=1}^{H}\big[\mathrm{Cov}_M(\pi^\star_h \,\|\, \widehat{\pi}_h) + \mathrm{Cov}^{\mathsf{act}}_{\eta,h}(\pi^\star \,\|\, \pi_{\mathsf{ref}})\big],$$

where we define

$$\begin{aligned}\mathrm{Cov}^{\mathsf{act}}_{\eta,h}(\pi^\star \,\|\, \pi_{\mathsf{ref}}) &:= \mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1}, x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{\pi^\star(x_h \mid x_{h-1})}{\pi_{\mathsf{ref}}(x_h \mid x_{h-1})} \ge \eta\right) \\ &= \mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1}, x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{V^\star(x_h)}{V^\star(x_{h-1})} \ge \eta\right).\end{aligned}$$

The quantity $\mathrm{Cov}^{\mathsf{act}}_{\eta,h}(\pi^\star \,\|\, \pi_{\mathsf{ref}})$ is the action-level analogue of Definition 3.4. Under Assumption 3.1, this additional term is 0 when $\eta \ge C_{\mathsf{act}}$, recovering Theorem 3.5.

Proof. We follow the notation of the proof of Theorem 3.2 (Appendix D.2). Recall that

$$M_h - M_{h-1} = \widehat{W}_{h-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{V^\star(x^i_h)}{\widehat{V}(\tilde{x}^i_{h-1})} - \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right),$$

where the tuples $(\tilde{x}^i_{h-1}, x^i_h)$ are i.i.d. given $\mathcal{H}_{h-1}$. By Proposition C.1, for any fixed parameter $R > 0$, we can bound

$$\begin{aligned}&\frac{1}{\widehat{W}_{h-1}}\,\mathbb{E}_{h-1}|M_h - M_{h-1}| \\ &\le \sqrt{\frac{R}{N}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\right]} + 2\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}, x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\mathbb{I}\left\{\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})} \ge R\right\}\right]\end{aligned}$$

By Lemma D.2, we have

$$\mathbb{E}\left[\widehat{W}_{h-1}\sqrt{\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]}\right] \le \sqrt{\mathbb{E}[\widehat{W}_{h-1}]\,\mathbb{E}[\widehat{W}_{h-1}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]]} \le \sqrt{\widehat{Z}_{h-1}\cdot Z},$$

and similarly, for any $\eta > 0$,

$$\begin{aligned}
&\mathbb{E}\left[\widehat{W}_{h-1}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1},x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\mathbb{I}\left\{\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\geq R\right\}\right]\right] \\
&= \mathbb{E}_{x_{h-1}\sim\pi_{h-1},x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[V^\star(x_h)\mathbb{I}\left\{\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\geq R\right\}\right] \\
&\leq \mathbb{E}_{x_{h-1}\sim\pi_{h-1},x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[V^\star(x_h)\mathbb{I}\left\{\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\geq \eta^{-1}R\right\}+V^\star(x_h)\mathbb{I}\left\{\frac{V^\star(x_h)}{V^\star(x_{h-1})}\geq\eta\right\}\right] \\
&= Z\cdot\left[\mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1}}\left(\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\geq\eta^{-1}R\right)+\mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1},x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{V^\star(x_h)}{V^\star(x_{h-1})}\geq\eta\right)\right],
\end{aligned}$$

where we use $\pi_{h-1}(x_{h-1})\cdot V^\star(x_{h-1}) = \pi^\star_{h-1}(x_{h-1})\cdot Z$ for $x_{h-1}\in\mathcal{X}_{h-1}$ and $\pi_{\mathsf{ref}}(x_h \mid x_{h-1})\cdot V^\star(x_h) = \pi^\star(x_h \mid x_{h-1})\cdot Z$. Therefore, we can choose $R = M\eta\frac{Z}{\widehat{Z}_{h-1}}$ to derive

$$\begin{aligned}
\frac{1}{Z}\,\mathbb{E}|M_h - M_{h-1}| \leq \sqrt{\frac{M\eta}{N}} &+ 2\mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1}}\left(\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\geq M\cdot\frac{Z}{\widehat{Z}_{h-1}}\right) \\
&+ 2\mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1},x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{V^\star(x_h)}{V^\star(x_{h-1})}\geq\eta\right).
\end{aligned}$$

Taking summation completes the proof. □

D.5 Proof of Theorem 3.6

We first prove the following lemma.

Lemma D.7. *Under Assumption 3.3 and Assumption 3.1, for any $\delta\in(0,1)$, if we run SMC (Algorithm 1) with $N = \Omega(H\mathsf{C}_\infty^4 C_{\mathsf{act}}^2\log(e/\delta))$ particles, then*

$$\mathbb{E}\left[\left(\widehat{W}_H - 2Z\right)_+\right]\leq\delta Z.$$

Proof. We follow the notation of the proof of Theorem 3.2 (Appendix D.2). Note that by Assumption 3.1, it holds that almost surely, for all h and i,

$$0\leq\frac{V^\star(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)} = \frac{V^\star(x_h^i)}{V^\star(\tilde{x}_{h-1}^i)}\cdot\frac{V^\star(\tilde{x}_{h-1}^i)}{\widehat{V}(\tilde{x}_{h-1}^i)}\leq C_{\mathsf{act}}\mathsf{C}_\infty.$$

Let $B = \mathsf{C}_\infty^2 C_{\mathsf{act}}$. Since $M_h = \widehat{W}_h\,\mathbb{E}_{\widehat{\nu}_h}[\rho]$, using Eq. (1) with $g=\rho$, we can write

$$\frac{M_h}{M_{h-1}} = \frac{\frac{1}{N}\sum_{i=1}^N\frac{V^\star(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)}}{\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]}.$$

Conditioned on $\mathcal{H}_{h-1}$, the numerator is the empirical average of N independent random variables whose mean is $\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]$. Moreover, by Assumptions 3.1 and 3.3, after normalizing by this mean, each summand lies in $[0,B]$ and has mean 1. More explicitly, define

$$Y_i := \frac{\frac{V^\star(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)}}{\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]},\qquad \bar{Y} := \frac{1}{N}\sum_{i=1}^N Y_i.$$

Then $M_h/M_{h-1} = \bar{Y}$, and conditioned on $\mathcal{H}_{h-1}$, the random variables Y_i are independent, have mean 1, and lie in $[0,B]$. Thus, for any $p\geq 1$, the inequality $\log u\leq u-1$ gives $u^p\leq\exp(p(u-1))$

for all $u \geq 0$, and hence Hoeffding's lemma gives

$$\mathbb{E}_{h-1}\left[\left(\frac{M_h}{M_{h-1}}\right)^p\right] = \mathbb{E}_{h-1}\left[\bar{Y}^p\right] \leq \mathbb{E}_{h-1}\left[\exp\left(p(\bar{Y}-1)\right)\right]$$
$$= \prod_{i=1}^{N} \mathbb{E}_{h-1}\left[\exp\left(\frac{p}{N}(Y_i - 1)\right)\right] \leq \exp\left(\frac{p^2B^2}{8N}\right).$$

Iterating over h gives

$$\mathbb{E}\left[\left(\frac{\widehat{W}_H}{Z}\right)^p\right] = \mathbb{E}\left[\left(\frac{M_H}{Z}\right)^p\right] \leq \exp\left(\frac{Hp^2B^2}{8N}\right).$$

Therefore, for any $p > 1$,

$$\frac{1}{Z}\,\mathbb{E}\left[\left(\widehat{W}_H - 2Z\right)_+\right] = \int_2^\infty \mathbb{P}\left(\widehat{W}_H \geq tZ\right) dt \leq \exp\left(\frac{Hp^2B^2}{8N}\right)\int_2^\infty t^{-p}\,dt.$$

Choosing $p = \Theta(\log(e/\delta))$ and using $N = \Omega(HB^2p)$ makes the final display at most δ, completing the proof. □

Proof of Theorem 3.6. In any given round, let $\widehat{p}$ be the sub-probability measure on $\mathcal{X}_H$ induced by the output of Algorithm 1 with Option 2 (when Algorithm 1 restarts we discard the output). Write $A = 2\mathsf{C}_\infty \widehat{V}(\bot)$ and $\gamma = Z/A$. Then, for any $X \subset \mathcal{X}_H$,

$$\widehat{p}(X) := \mathbb{E}\left[\min\left\{1, \frac{\widehat{W}_H}{A}\right\} \cdot \widehat{\nu}_H(X)\right] \leq \frac{Z \cdot \pi^\star_H(X)}{A} = \gamma\pi^\star_H(X),$$

where we have used Lemma D.2. Moreover, for $\alpha \in (0, 1/2)$ to be chosen below, Lemma D.7 gives $\mathbb{E}[(\widehat{W}_H - 2Z)_+] \leq \alpha Z$ as long as $N \geq \Omega(H\mathsf{C}^4_\infty C^2_{\mathsf{act}} \log(eH/\alpha))$. Since $V^\star(\bot) = Z$, Assumption 3.3 gives $Z/\mathsf{C}_\infty \leq \widehat{V}(\bot) \leq \mathsf{C}_\infty Z$, and in particular $A = 2\mathsf{C}_\infty\widehat{V}(\bot) \geq 2Z$. Thus, the inequality $\min\{1, y\} \geq y - (y-1)_+$ gives

$$\widehat{p}(X) \geq \mathbb{E}\left[\frac{\widehat{W}_H}{A}\widehat{\nu}_H(X)\right] - \mathbb{E}\left[\left(\frac{\widehat{W}_H}{A} - 1\right)_+\right]$$
$$\geq \gamma\pi^\star_H(X) - \frac{\mathbb{E}[(\widehat{W}_H - 2Z)_+]}{A} \geq \gamma(\pi^\star_H(X) - \alpha).$$

In particular, the success probability $s := \widehat{p}(\mathcal{X}_H)$ satisfies $\gamma(1-\alpha) \leq s \leq \gamma$. Let μ denote the law of the output conditioned on success. Then, for every $X \subset \mathcal{X}_H$,

$$\pi^\star_H(X) - \alpha \leq \mu(X) \leq \frac{\pi^\star_H(X)}{1-\alpha} \leq \pi^\star_H(X) + 2\alpha,$$

where the final inequality uses $\alpha \leq 1/2$. Taking α to be a sufficiently small constant multiple of δ gives $D_{\mathsf{TV}}(\mu, \pi^\star_H) \leq \delta$. Finally, using the guarantee of rejection sampling, we know the number of outer loop steps is bounded by $O(\mathsf{C}^2_\infty \log(1/\delta))$ by noting that $\widehat{V}(\bot)/Z \leq \mathsf{C}_\infty$. □

D.6 Lower Bound for SMC with Exact Value Function

In the following proposition, we show that the horizon dependence of SMC when $\widehat{V} = V^\star$ is not an artifact of our analysis.

Proposition D.8. *There exists a problem instance where $\widehat{V} = V^\star$ and Assumption 3.1 holds with $C_{\mathsf{act}} = 2$, however SMC with N particles must incur*

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \geq c\min\left\{1, \frac{\sqrt{H}}{N}\right\}.$$

Proof. We construct a problem instance as follows. Fix $\lambda \in (0,1)$. We consider $\mathcal{X}_h = \{0,1\}^h$, and define the transition kernel P and value function $\widehat{V} = V^\star$ as follows:

- For any x, let $\pi_{\mathsf{ref}}(\cdot \mid x) = \mathrm{unif}(\{(x,0),(x,1)\})$.
- Let $\widehat{V} = V^\star : \mathcal{X} \to \mathbb{R}_{\geq 0}$ be defined recursively as
$$V^\star(\bot) = 1, \qquad V^\star(x,0) = (1-\lambda)V^\star(x), \quad V^\star(x,1) = (1+\lambda)V^\star(x).$$
We take the terminal reward to be $r^\star(x_H) = V^\star(x_H)$. Since $V^\star(x) = \mathbb{E}_{x' \sim \pi_{\mathsf{ref}}(\cdot|x)}[V^\star(x')]$ by construction, this recursive definition agrees with the value function induced by $r^\star$.
- Then, it holds that $C_{\mathsf{act}} \leq 1+\lambda$, and $\pi^\star_H = \mathrm{Ber}(p)^{\otimes H}$, where $p = \frac{1+\lambda}{2}$.

Under this construction, we can rewrite SMC as follows (where $\widehat{\nu}_0 = \delta_\bot$).

- For $h \geq 1$, sample N particles $\tilde{x}^1_{h-1}, \ldots, \tilde{x}^N_{h-1} \overset{\text{i.i.d.}}{\sim} \widehat{\nu}_{h-1}(\cdot)$ and generate $a^i_h \sim \mathrm{unif}(\{0,1\})$ independently.
- Assign weights $w^i_h = 1 + \lambda(2a^i_h - 1)$ to particle $x^i_h = (\tilde{x}^i_{h-1}, a^i_h)$.
- Define weighted empirical measure $\widehat{\nu}_h := \frac{1}{W_h}\sum_{i=1}^N w^i_h \delta_{x^i_h}$, where $W_h = N(1-\lambda) + 2\lambda\sum_{i=1}^N a^i_h$.

We make the following claims, proven below.

Claim D.9. *It holds that $\mathbb{E}[\widehat{\nu}_H] = \mathrm{Ber}(p_N)^{\otimes H}$, where p_N is defined in Eq. (4).*

Claim D.10. *It holds that $p_N \leq \frac{1+\lambda}{2} - \frac{c_\lambda}{N}$, where $c_\lambda = \frac{\lambda(1-\lambda^2)}{2}$.*

Now, we can lower bound
$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) = D_{\mathsf{TV}}(\mathrm{Bin}(H,p), \mathrm{Bin}(H,p_N)) \geq c\min\{1, \sqrt{H}|p - p_N|\},$$
where $c > 0$ is an absolute constant. Choosing $\lambda = 1/2$ completes the proof of Proposition D.8. □

We prove the auxiliary claims:

Proof of Claim D.9. By definition, for any $x \in \mathcal{X}_{h-1}$ and $a \in \{0,1\}$,
$$\widehat{\nu}_h(x,a) = \frac{1}{W_h}\sum_{i=1}^N w^i_h \mathbb{I}\{\tilde{x}^i_{h-1} = x, a^i_h = a\}.$$
Note that $\{a^i_h\}$ and $\{\tilde{x}^i_{h-1}\}$ are independent, so we can calculate
$$\begin{aligned}\mathbb{E}[\widehat{\nu}_h(x,1)] &= \mathbb{E}\left[\frac{1}{W_h}\sum_{i=1}^N w^i_h \mathbb{I}\{a^i_h = 1\} \cdot \mathbb{I}\{\tilde{x}^i_{h-1} = x\}\right] \\ &= \mathbb{E}\left[\widehat{\nu}_{h-1}(x) \cdot \frac{(1+\lambda)\sum_{i=1}^N a^i_h}{N(1-\lambda) + 2\lambda\sum_{i=1}^N a^i_h}\right].\end{aligned}$$
Note that $A_h = \sum_{i=1}^N a^i_h \sim \mathrm{Bin}(N, \frac{1}{2})$ is independent of $\widehat{\nu}_{h-1}$. Then, we know
$$\mathbb{E}[\widehat{\nu}_h(x,1)] = p_N\,\mathbb{E}[\widehat{\nu}_{h-1}(x)], \qquad p_N = (1+\lambda)\,\mathbb{E}_{A \sim \mathrm{Bin}(N,\frac{1}{2})}\left[\frac{A}{N(1-\lambda) + 2\lambda A}\right]. \tag{4}$$
Applying this equation recursively gives the desired result. □

Proof of Claim D.10. We let $Y = \frac{2A}{N} - 1$ so that Y is symmetrically distributed on $[-1,1]$. Then, with $p = \frac{1+\lambda}{2}$, we can write
$$p_N = \frac{1+\lambda}{2}\,\mathbb{E}\left[\frac{1+Y}{1+\lambda Y}\right].$$

Algorithm 2 Sequential Monte Carlo with Rejection Sampling

Input: Transition kernel $\pi_{\mathtt{ref}}$, value function $\widehat{V}$.
Parameter: initial sample size $N \geq 1$, parameter $\eta \geq 1$.
Initialize $\mathcal{S}_0$ to be N copies of $\perp$.
for $h = 1, \cdots, H$ **do**
  Initialize a multi-set $\mathcal{S}_h = \{\}$.
  while $|\mathcal{S}_h| < N$ **do**
    Sample $x_{h-1} \sim \mathrm{Unif}(\mathcal{S}_{h-1})$.
    Sample $x_h \sim \pi_{\mathtt{ref},h-1}(\cdot \mid x_{h-1})$.
    With probability $\frac{\widehat{V}(x_h)}{\eta \widehat{V}(x_{h-1})}$, update $\mathcal{S}_h \leftarrow \mathcal{S}_h \cup \{x_h\}$.
Output: $x \sim \widehat{\nu}_H = \mathsf{Unif}(\mathcal{S}_H)$.

Therefore,

$$p - p_N = \frac{1-\lambda^2}{2}\,\mathbb{E}\left[\frac{-Y}{1+\lambda Y}\right] = \frac{1-\lambda^2}{2\lambda}\left(\mathbb{E}\left[\frac{1}{1+\lambda Y}\right] - 1\right).$$

Since $\lambda \in (0,1)$, the geometric series expansion is valid; using that Y is symmetric and $\mathbb{E}\,Y^2 = \frac{1}{N}$, we have

$$\mathbb{E}\left[\frac{1}{1+\lambda Y}\right] = \sum_{k\geq 0} \lambda^{2k}\,\mathbb{E}\big[Y^{2k}\big] \geq 1 + \frac{\lambda^2}{N}.$$

It follows that

$$p_N \leq \frac{1+\lambda}{2} - \frac{\lambda(1-\lambda^2)}{2N}.$$

□

E Proofs for SMC-RS

Analogously to Appendix D, we introduce the following quantity for Algorithm 2:

$$W_h := \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}, x_h\sim\pi_{\mathtt{ref}}(\cdot|x_{h-1})}\left[\frac{\widehat{V}(x_h)}{\widehat{V}(x_{h-1})}\right], \qquad \widehat{W}_h := \widehat{V}(\perp)\prod_{\ell=1}^{h} W_\ell. \tag{5}$$

We define $\widehat{\nu}_h := \mathrm{Unif}(\mathcal{S}_h)$ and $\mathcal{H}_h = \{\mathcal{S}_1, \cdots, \mathcal{S}_h\}$ to be the full history of the execution of SMC-RS (Algorithm 2) up to the end of iteration h. We write $\mathbb{E}_h[\cdot]$ to be the conditional expectation with respect to $\mathcal{H}_h$.

Lemma E.1. *For step $h \in [H]$ of SMC-RS (Algorithm 2), given $\mathcal{H}_{h-1}$, the set $\mathcal{S}_h$ consists of i.i.d. samples from the distribution $\widehat{\mu}_h$ given by*

$$\widehat{\mu}_h(x_h) = \frac{1}{W_h}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\pi_{\mathtt{ref}}(x_h \mid x_{h-1})\frac{\widehat{V}(x_h)}{\widehat{V}(x_{h-1})}\right], \qquad \forall x_h \in \mathcal{X}_h.$$

Hence, for any $g_h : \mathcal{X}_h \to \mathbb{R}$, it holds that

$$\mathbb{E}\Big[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g_h]\Big] = \mathbb{E}\Big[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\mu}_h}[g_h]\Big] = \mathbb{E}_{x_h\sim\pi_h}\Big[\widehat{V}(x_h) g_h(x_h)\Big]. \tag{6}$$

Proof of Lemma E.1. The first claim follows immediately from the analysis of rejection sampling: For any $x_h \in \mathcal{X}_h$, by Bayes' rule, we have

$$\mathbb{P}(x_h \mid \text{accept}) \propto \mathbb{P}(\text{accept} \mid x_h) \cdot \mathbb{P}(x_h) \propto \widehat{\mu}_h(x_h).$$

To prove the second claim, we prove the following result, from which Eq. (6) follows inductively:

$$\mathbb{E}_{h-1}[W_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g_h]] = \mathbb{E}_{h-1}[W_h \cdot \mathbb{E}_{\widehat{\mu}_h}[g_h]] = \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}, x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{\widehat{V}(x_h)g_h(x_h)}{\widehat{V}(x_{h-1})}\right].$$

Since $\widehat{W}_0 = \widehat{V}(\perp)$ and $\widehat{\nu}_0 = \delta_\perp$, Eq. (6) follows by induction on h. □

In particular, $\mathbb{E}[\widehat{W}_h] = \widehat{Z}_h$ for $h \geq 1$ and

$$\mathbb{E}\left[\widehat{W}_H \widehat{\nu}_H(x)\right] = Z \cdot \pi^\star_H(x), \qquad \forall x \in \mathcal{X}_H.$$

Further, we define $\rho(x) = \frac{V^\star(x)}{\widehat{V}(x)}$ and

$$M_h = \widehat{W}_h \cdot \mathbb{E}_{x_h\sim\widehat{\nu}_h}[\rho(x_h)].$$

Then Lemma E.1 with $g(x) = \rho(x)$ implies (M_h) is a martingale with $M_0 = Z$ and $M_H = \widehat{W}_H$. Therefore, similar to Lemma D.3, we have the following upper bound.

Lemma E.2. *For Algorithm 2, it holds that*

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq \frac{1}{2Z}\,\mathbb{E}|\widehat{W}_H - Z| \leq \frac{1}{2Z}\sum_{h=1}^{H}\mathbb{E}|M_h - M_{h-1}|.$$

E.1 Proof of Theorem 3.8

Note that $M_h - M_{h-1} = \widehat{W}_h(\mathbb{E}_{\widehat{\nu}_h}[\rho] - \mathbb{E}_{\widehat{\mu}_h}[\rho])$, and $\widehat{\nu}_h = \mathrm{unif}(\mathcal{S}_h)$, where $\mathcal{S}_h$ consists of i.i.d samples from $\widehat{\mu}_h$. Then, we can bound

$$\mathbb{E}_{h-1}|\mathbb{E}_{\widehat{\nu}_h}[\rho_h] - \mathbb{E}_{\widehat{\mu}_h}[\rho_h]| \leq \sqrt{\mathbb{E}_{h-1}\left|\frac{1}{N}\sum_{x_h\in\mathcal{S}_h}\rho(x_h) - \mathbb{E}_{\widehat{\mu}_h}[\rho]\right|^2} = \sqrt{\frac{\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)^2] - (\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2}{N}},$$

and hence

$$\begin{aligned}
\sqrt{N}\,\mathbb{E}|M_h - M_{h-1}| &\leq \mathbb{E}\left[\widehat{W}_h\sqrt{\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)^2] - (\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2}\right] \\
&\leq \sqrt{\mathbb{E}[\widehat{W}_h]\left(\mathbb{E}\left[\widehat{W}_h\,\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)^2]\right] - \mathbb{E}\left[\widehat{W}_h(\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2\right]\right)} \\
&\leq \sqrt{\widehat{Z}_h\,\mathbb{E}_{x_h\sim\pi_h}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_h)}\right] - Z^2} \\
&= Z\sqrt{\mathbb{E}_{x_h\sim\pi^\star_h}\left[\frac{\pi^\star(x_h)}{\widehat{\pi}(x_h)}\right] - 1} = Z\sqrt{D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h)},
\end{aligned}$$

where the third line uses $\mathbb{E}[\widehat{W}_h]\,\mathbb{E}\left[\widehat{W}_h(\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2\right] \geq \left(\mathbb{E}[\widehat{W}_h\,\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)]]\right)^2 = Z^2$ (Cauchy inequality and Lemma E.1), and the last line follows from the definition of $\pi^\star$, $\widehat{\pi}$, and χ^2-divergence. Taking summation completes the proof of the total variation bound.

It remains to prove the expected running time bound. Let T_h be the number of proposals made by Algorithm 2 during iteration h. Conditioned on $\mathcal{H}_{h-1}$, each proposal is accepted with probability

$$p_h = \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1},\, x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{\widehat{V}(x_h)}{\eta\widehat{V}(x_{h-1})}\right] = \frac{W_h}{\eta}.$$

By Assumption 3.7 and the condition $\eta \geq \widehat{C}_{\mathsf{act}}$, this is a valid acceptance probability, and the lower-bound part of Assumption 3.7 gives $W_h \geq \widehat{C}^{-1}_{\mathsf{act}}$. Since T_h is negative binomial with N successes and success probability p_h, we have

$$\mathbb{E}_{h-1}[T_h] = \frac{N}{p_h} = \frac{N\eta}{W_h} \leq N\eta\widehat{C}_{\mathsf{act}}.$$

Summing over $h \in [H]$ gives $\mathbb{E}[\sum_{h=1}^{H} T_h] \leq NH\eta\widehat{C}_{\mathsf{act}}$, which is the claimed expected time bound. □

E.2 Analysis of SMC-RS with Heavy-Tailed Errors

Parallel to SMC, we show that SMC-RS also enjoys an upper bound that depends only on coverage (Definition 3.4).

Theorem E.3 (SMC-RS convergence under coverage)**.** *Under Assumption 3.7, suppose that SMC-RS (Algorithm 2) is instantiated with $\eta \geq \widehat{C}_{\mathsf{act}}$. It holds that for any $M \geq 1$, Algorithm 2 achieves*

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq H\sqrt{\frac{M}{N}} + \sum_{h=1}^{H} \mathrm{Cov}_M(\pi^\star_h \,\|\, \widehat{\pi}_h).$$

The convergence rate of Theorem E.3 improves upon that of SMC (Theorem 3.5) by a factor of C_{act}, though the same factor reappears in the time complexity. The expected time complexity bound matches Theorem 3.8.

Proof. We follow the proof of Theorem 3.8 (Appendix E.1). By Proposition C.1, it holds that

$$\begin{aligned}\mathbb{E}_{h-1}|\mathbb{E}_{\widehat{\nu}_h}[\rho_h] - \mathbb{E}_{\widehat{\mu}_h}[\rho_h]| &= \mathbb{E}_{h-1}\left|\frac{1}{N}\sum_{x_h \in \mathcal{S}_h} \rho(x_h) - \mathbb{E}_{\widehat{\mu}_h}[\rho]\right| \\ &\leq \sqrt{\frac{R}{N}\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]} + 2R\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)\mathbb{I}\{\rho(x_h) \geq R\}],\end{aligned}$$

where we use $\mathcal{S}_h$ consists of N i.i.d samples from $\widehat{\mu}_h$. Therefore, we can bound

$$\mathbb{E}|M_h - M_{h-1}| \leq 2\sqrt{\frac{R}{N}}\,\mathbb{E}\left[\widehat{W}_h\sqrt{\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]}\right] + 2\,\mathbb{E}\left[\widehat{W}_h\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)\mathbb{I}\{\rho(x_h) \geq R\}]\right]$$

Further, by Lemma E.1, we have

$$\mathbb{E}\left[\widehat{W}_h\sqrt{\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]}\right] \leq \sqrt{\mathbb{E}[\widehat{W}_h]\,\mathbb{E}[\widehat{W}_h\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]]} \leq \sqrt{\widehat{Z}_h \cdot Z},$$

and

$$\mathbb{E}\left[\widehat{W}_h \cdot \mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)\mathbb{I}\{\rho(x_h) \geq R\}]\right] = Z \cdot \mathbb{P}_{x\sim\pi^\star_h}(\rho(x_h) \geq R).$$

Choosing $R = M\frac{Z}{\widehat{Z}_h}$ gives

$$\mathbb{E}|M_h - M_{h-1}| \leq 2Z\sqrt{\frac{M}{N}} + 2Z\mathrm{Cov}_M(\pi^\star_h \,\|\, \widehat{\pi}_h).$$

Taking summation completes the proof. □

E.3 Faster Convergence for SMC-RS

In this section, we present and analyze a modification of SMC-RS, which we call SMC-RS with restart (Algorithm 3). Under the following assumption, we prove that Algorithm 3 achieves exponential convergence (analogous to Theorem 3.6 for SMC).

Assumption E.4 (Finite action support)**.** *There are at most η values of x_{h+1} for each x_h so that $\pi_{\mathsf{ref}}(x_{h+1} \mid x_h) > 0$, and these are efficiently enumerable.*

We prove the following guarantee:

Theorem E.5 (Fast-rate SMC-RS under L^∞ condition)**.** *Suppose that Assumption 3.3 and Assumption E.4 hold, and we denote $\epsilon := \mathsf{C}_\infty - 1$. Further, suppose that we have $Z \leq \widetilde{Z}_1 \leq JZ$. Then Algorithm 3 samples from a distribution μ that satisfies*

$$D_{\mathsf{TV}}(\mu, \pi^\star_H) \leq \delta,$$

as long as $N \geq \Omega(H\mathsf{C}^2_\infty\epsilon^2\log(JH/\delta))$, using a number of outer loop steps that is bounded by $O(J^2\log(1/\delta))$.

Algorithm 3 SMC-RS with restart

Input: Transition kernel $\pi_{\mathtt{ref}}$, value function $\widehat{V}$, estimate $\widetilde{Z}_1$ of $\widehat{Z}_1$.
Parameter: initial sample size $N \geq 1$.
Initialize $\mathcal{S}_0$ to be N copies of $\perp$.
for $h = 1, \cdots, H$ **do**
For each $x_{h-1} \in \mathcal{S}_{h-1}$, compute $\widetilde{V}(x_{h-1}) = \mathbb{E}_{\pi_{\mathtt{ref}}}[\widehat{V}(x_h) \mid x_{h-1}]$ and the distribution $\widehat{\pi}_h(\cdot \mid x_{h-1}) \propto \pi_{\mathtt{ref}}(\cdot \mid x_{h-1}) \cdot \widehat{V}(\cdot)$, which takes time $O(\eta)$ using Assumption E.4.
Define $W_h = \frac{1}{N} \sum_{x_{h-1} \in \mathcal{S}_{h-1}} \frac{\widetilde{V}(x_{h-1})}{\widehat{V}(x_{h-1})}$.
Define the weighted empirical measure $\widetilde{\mu}_{h-1} = \frac{1}{NW_h} \sum_{x_{h-1} \in \mathcal{S}_{h-1}} \frac{\widetilde{V}(x_{h-1})}{\widehat{V}(x_{h-1})} \delta_{x_{h-1}}$.
Generate $\tilde{x}^i_{h-1} \sim \widetilde{\mu}_{h-1}, x^i_h \sim \widehat{\pi}_h(\cdot \mid \tilde{x}^i_{h-1})$ independently for $i \in [N]$.
Define $\mathcal{S}_h = \{x^i_h\}_{i \in [N]}$ and $\widehat{\nu}_h = \mathsf{Unif}(\mathcal{S}_h)$.
Define $\widehat{W}_H = \widehat{V}(\perp) \prod_{h=1}^H W_h$.
Output: With probability $\min\left\{\frac{\widehat{W}_H}{\widetilde{Z}_1}, 1\right\}$, output $x \sim \widehat{\nu}_H$, and otherwise restart the algorithm.

Under Assumption E.4, we can always choose $\widetilde{Z}_1 = \mathsf{C}_\infty \widehat{Z}_1$, so that $Z \leq \widetilde{Z}_1 \leq \mathsf{C}_\infty^2 Z$.[7]

Similar to the proof of Theorem 3.6, we first state and prove the following lemma (as an analogue to Lemma D.7). Theorem E.5 then follows from the argument in Appendix D.5.

Lemma E.6. *Under the conditions of Theorem E.5, for any $\delta \in (0,1)$, if we run SMC-RS with restart (Algorithm 3) with $N = \Omega(H\mathsf{C}_\infty^2 \epsilon^2 \log(e/\delta))$ particles, then*

$$\mathbb{E}\left[\left(\widehat{W}_H - 2Z\right)_+\right] \leq \delta Z.$$

Proof of Lemma E.6. We follow the proof of Lemma D.7. Moreover, note that the definition of W_h in Algorithm 3 concurs with that in Eq. (5); further, let $\widehat{W}_h$ be defined as in Eq. (5). Letting $M_h = \widehat{W}_h \mathbb{E}_{\widehat{\nu}_h}[\rho]$, so that $M_{h-1} = \widehat{W}_h \mathbb{E}_{\widehat{\mu}_h}[\rho]$ (see Lemma E.1), we can write

$$\frac{M_h}{M_{h-1}} = \frac{\frac{1}{N}\sum_{i=1}^N \rho(x^i_h)}{\mathbb{E}_{\widehat{\mu}_h}[\rho]},$$

where $x^1_h, \cdots, x^N_h$ are N i.i.d. samples from $\widehat{\mu}_h$ given $\mathcal{H}_{h-1}$. Let $B = 2\mathsf{C}_\infty \epsilon$. By Assumption 3.3, $\rho(x_h) \in [\mathsf{C}_\infty^{-1}, \mathsf{C}_\infty]$ for all $x_h \in \mathcal{X}_h$, and hence, after normalizing by $\mathbb{E}_{\widehat{\mu}_h}[\rho]$, each summand has mean 1 and lies in an interval of length at most B. More explicitly, define

$$Y_i := \frac{\rho(x^i_h)}{\mathbb{E}_{\widehat{\mu}_h}[\rho]}, \qquad \bar{Y} := \frac{1}{N}\sum_{i=1}^N Y_i.$$

Then $M_h/M_{h-1} = \bar{Y}$, and conditioned on $\mathcal{H}_{h-1}$, the random variables Y_i are independent and have mean 1. Thus, for any $p \geq 1$, the inequality $\log u \leq u - 1$ gives $u^p \leq \exp(p(u-1))$ for all $u \geq 0$, and hence Hoeffding's lemma gives

$$\begin{aligned}\mathbb{E}_{h-1}\left[\left(\frac{M_h}{M_{h-1}}\right)^p\right] = \mathbb{E}_{h-1}\left[\bar{Y}^p\right] &\leq \mathbb{E}_{h-1}\left[\exp\left(p(\bar{Y}-1)\right)\right] \\ &= \prod_{i=1}^N \mathbb{E}_{h-1}\left[\exp\left(\frac{p}{N}(Y_i - 1)\right)\right] \leq \exp\left(\frac{p^2B^2}{8N}\right).\end{aligned}$$

Iterating over h gives

$$\mathbb{E}\left[\left(\frac{\widehat{W}_H}{Z}\right)^p\right] = \mathbb{E}\left[\left(\frac{M_H}{Z}\right)^p\right] \leq \exp\left(\frac{Hp^2B^2}{8N}\right).$$

[7] Further, we can run Algorithm 3 1 time solely for estimating Z, with choice $\widetilde{Z}_1 = 2\widehat{W}_H$ satisfying $Z \leq \widetilde{Z} \leq 4Z$ with high probability.

Therefore, for any $p > 1$,

$$\frac{1}{Z}\mathbb{E}\left[\left(\widehat{W}_H - 2Z\right)_+\right] = \int_2^\infty \mathbb{P}\left(\widehat{W}_H \geq tZ\right) dt \leq \exp\left(\frac{Hp^2B^2}{8N}\right)\int_2^\infty t^{-p}\, dt.$$

Choosing $p = \Theta(\log(e/\delta))$ and using $N = \Omega(HB^2p)$ makes the final display at most δ, completing the proof. □

F A Unified Perspective on Backtracking and Multi-Particle Methods

Our main theoretical results demonstrated that particle filtering methods like SMC can achieve strong guarantees under the same assumptions on $\widehat{V}$ as the backtracking-based method VGB introduced by [43] (which in turn derives from the classical Sinclar-Jerrum random walk [46]), while improving parallel runtime from $\widetilde{O}(H^2)$ to $\widetilde{O}(H)$. In this section, we show that the connection goes deeper: for *any* problem instance, the execution of VGB can be *coupled* with the execution of SMC-IND, a variant of SMC in which the children of each particle are sampled independently.

We defer formal definitions until after stating the result (Theorem F.1), but we make several remarks for context. Henceforth, we specify to the autoregressive setting (Remark 2.1) in which VGB was defined [43, 46]. Recall that VGB implements a single-particle random walk in which the allowed transitions can not only add a token to the current particle (i.e. $a_{1:h} \mapsto a_{1:h+1}$) but also backtrack (i.e. $a_{1:h} \mapsto a_{1:h-1}$). For technical reasons, we augment the state space of the random walk with a state $\mathfrak{s}$ that is only adjacent to the empty string $\perp$. See Appendices F.1 and F.2 for the formal definitions of SMC-IND and the VGB random walk.

Theorem F.1. *Let $\mathcal{S}_{\mathsf{DMC}}$ be the (random) multiset of particles visited by SMC-IND with N initial particles. Let $\mathcal{S}_{\mathsf{VGB}}$ be the (random) multiset of particles visited by VGB (initialized at $\mathfrak{s}$) before the $N+1$-st visit to $\mathfrak{s}$. Then there is a coupling so that $\mathcal{S}_{\mathsf{DMC}} = \mathcal{S}_{\mathsf{VGB}} \setminus \{\mathfrak{s}\}$ (with multiplicity) almost surely.*

Intuitively, each particle $a_{1:h}$ visited by SMC-IND corresponds to an interval of time in which the VGB random walk started at $a_{1:h}$, ended at $a_{1:h}$, and never visited the parent node $a_{1:h-1}$ (where we take the parent of $\perp$ to be $\mathfrak{s}$). This demonstrates that the paradigm of backtracking to tame error amplification [43] is, in a sense, equivalent to a form of particle filtering.

Outline of appendix. In Appendix F.1 we introduce the VGB random walk (slightly modified for technical reasons). In Appendix F.2 we introduce the SMC-IND stochastic process. In Appendix F.3 we prove Theorem F.1.

F.1 Background on VGB

The (non-lazy) VGB random walk [43] is defined on state space $\mathcal{X} = \mathcal{X}_0 \sqcup \cdots \sqcup \mathcal{X}_H$ where $\mathcal{A}$ is the token space and $\mathcal{X}_h := \mathcal{A}^h$. For technical convenience (specifically, to avoid casework at step 0 when coupling VGB with SMC-IND), we introduce an auxiliary state $\mathfrak{s}$ that is not in the original definition. We define the transition kernel $\mathbb{P}_{\mathsf{VGB}}$ as follows. First, $\mathfrak{s}$ always transitions to $\perp$:

$$\mathbb{P}_{\mathsf{VGB}}(\perp \mid \mathfrak{s}) = 1.$$

Second, any $a_{1:h} \in \mathcal{A}^h$ with $0 \leq h < H$ can transition to either a "child" $a_{1:h+1}$ or the "parent" $a_{1:h-1}$:

$$\mathbb{P}_{\mathsf{VGB}}(a_{1:h+1} \mid a_{1:h}) = \frac{\pi_{\mathsf{ref}}(a_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h+1})}{\widehat{V}(a_{1:h}) + \sum_{a'_{h+1}} \pi_{\mathsf{ref}}(a'_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h}, a'_{h+1})},$$

$$\mathbb{P}_{\mathsf{VGB}}(a_{1:h-1} \mid a_{1:h}) = \frac{\widehat{V}(a_{1:h})}{\widehat{V}(a_{1:h}) + \sum_{a'_{h+1}} \pi_{\mathsf{ref}}(a'_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h}, a'_{h+1})},$$

with the convention that $a_{1:0} = \perp$ (i.e. the empty string) and $a_{1:-1} = \mathfrak{s}$. Third, any $a_{1:H} \in \mathcal{A}^H$ must transition to its parent:

$$\mathbb{P}_{\mathsf{VGB}}(a_{1:H-1} \mid a_{1:H}) = 1.$$

The transition kernel of the original VGB random walk [43] is identical conditioned on never visiting $\mathfrak{s}$. Moreover, since $\mathfrak{s}$ is only reachable from $\perp$ (and only transitions to $\perp$), our modification only affects the execution of VGB by stochastically inserting copies of the sequence $(\mathfrak{s}, \perp)$ into VGB's random trajectory.

For notational convenience, it is useful to define a transition kernel $\mathbb{P}_{\mathsf{down}}$ as the VGB transition kernel conditioned on "downward" moves, i.e.

$$\mathbb{P}_{\mathsf{down}}(a_{1:h+1} \mid a_{1:h}) = \frac{\pi_{\mathtt{ref}}(a_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h+1})}{\sum_{a'_{h+1}} \pi_{\mathtt{ref}}(a'_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h}, a'_{h+1})}.$$

F.2 Definition of SMC-IND

In Algorithm 4, we describe how SMC-IND stochastically constructs multisets of particles $\mathcal{S}_0, \ldots, \mathcal{S}_H$ for a given problem instance. Note that there is no specified output mechanism, since the coupling between SMC-IND and VGB will apply to the entire multiset of particles.

Algorithm 4 Sequential Monte Carlo with Independent Sampling (SMC-IND)

Input: Autoregressive language model $\pi_{\mathtt{ref}}$, value function $\widehat{V}$.
Parameter: initial sample size $N \geq 1$.
Initialize $\mathcal{S}_0$ to be N copies of $\perp$.
for $h = 1, \cdots, H$ **do**
  Initialize a multi-set $\mathcal{S}_h = \{\}$.
  for $a_{1:h-1} \in \mathcal{S}_{h-1}$ **do**
    Sample $D \sim \mathrm{Geom}(\mathbb{P}_{\mathsf{VGB}}(a_{1:h-2} \mid a_{1:h-1}))$ (with convention $a_{1:-1} = \mathfrak{s}$; see Appendix F.1)
    for $1 \leq i \leq D$ **do**
      Sample $a_{1:h} \sim \mathbb{P}_{\mathsf{down}}(\cdot \mid a_{1:h-1})$.
      Update $\mathcal{S}_h \leftarrow \mathcal{S}_h \cup \{a_{1:h}\}$.

F.3 Coupling VGB with SMC-IND

We interpret the executions of SMC-IND and VGB as stochastically constructing forests of particles. We then show that these forests are actually the same (up to occurrences of $\mathfrak{s}$) when the randomness is appropriately coupled.

Infinite-particle SMC-IND process. The execution of SMC-IND with $N = \infty$ naturally defines a rooted forest $\mathcal{F}_{\mathsf{DMC}}$ where each node is labelled by a particle in $\mathcal{S}_0 \sqcup \cdots \sqcup \mathcal{S}_H$. The roots are the particles in $\mathcal{S}_0$. For any $h > 0$, the parent of a particle in $\mathcal{S}_h$ is the particle in $\mathcal{S}_{h-1}$ that birthed it. Note that there may be multiple (or even infinitely many) occurrences of any particle $a_{1:h}$ in $\mathcal{S}_h$, so $\mathcal{F}_{\mathsf{DMC}}$ may contain multiple nodes with the same label. The parent of a node labelled $a_{1:h}$ always has label $a_{1:h-1}$.

Infinite-time VGB process. We next show how an execution of VGB in the autoregressive tree $\mathcal{T}$ can be interpreted as a rooted tree $\mathcal{F}_{\mathsf{VGB}}$, similar to above. Fix an infinite trajectory $(x_0, x_1, x_2, x_3, \ldots)$ of VGB with $x_0 = \mathfrak{s}$. Suppose that $\mathfrak{s}$ occurs infinitely often in the trajectory. The root node of $\mathcal{F}_{\mathsf{VGB}}$ is indexed by $\mathbb{N}$, and labeled $\mathfrak{s}$. Every other node is indexed by an interval $[t, t'] \subset \mathbb{N}$ with the property that $x_{t-1} = x_{t'+1}$ is the parent of $x_t = x_{t'}$ in $\mathcal{T}$, and this parent does not occur within $\{x_t, x_{t+1}, \ldots, x_{t'}\}$. This node is labeled with x_t, and its parent in $\mathcal{F}_{\mathsf{VGB}}$ is the unique node with label x_{t-1} whose interval contains $[t, t']$.

Lemma F.2. *Suppose that VGB almost surely visits $\mathfrak{s}$ infinitely often. There is a coupling of the randomness of $\mathcal{F}_{\mathsf{DMC}}$ and $\mathcal{F}_{\mathsf{VGB}}$ such that $\mathcal{F}_{\mathsf{DMC}}$ is almost surely equal to the forest $\mathcal{F}_{\mathsf{VGB}} \setminus \{\mathfrak{s}\}$.*

Proof. We construct the coupling layer by layer. By the lemma assumption, it is clear that the first layer of $\mathcal{F}_{\mathsf{DMC}}$ is almost surely identical to that of $\mathcal{F}_{\mathsf{VGB}}$. Now fix $h \geq 1$ and condition on the forest structure and labels x_t (but not the indices $[t, t']$ themselves) of the first h layers of $\mathcal{F}_{\mathsf{VGB}}$. Fix any node ι with label $a_{1:h-1}$.

Consider the law of the VGB random walk between these two visits to $a_{1:h-1}$. The walk is restricted to the subtree rooted at $a_{1:h-1}$, but since the walk is Markovian, its law is otherwise unchanged by the conditioning (and it is also independent of the children of other nodes $\iota' \neq \iota$ at layer h). Let $p = \mathbb{P}(a_{1:h-2} \mid a_{1:h-1})$ (with the convention $a_{1:h-2} = \mathfrak{s}$ if $h = 1$). The number of children D' of ι is the number of times that the walk goes to a child of $a_{1:h-1}$ in $\mathcal{T}$ before it returns to $a_{1:h-2}$. Thus, $\mathbb{P}[D' = k] = (1-p)^k p$ for any $k \in \{0, 1, 2, \dots\}$. That is, D' is a geometric random variable with success probability p. Moreover, the label of each child of ι is independent and distributed according to $\mathbb{P}_{\mathsf{down}}(\cdot \mid a_{1:h-1})$. Thus, the coupling can be extended to the children of ι.

□

Proof of Theorem F.1. The first N trees of $\mathcal{F}_{\mathsf{DMC}}$ can be coupled to the forest induced by SMC-IND with N initial particles. Similarly, the first N trees of $\mathcal{F}_{\mathsf{VGB}}$ correspond exactly to the prefix of the trajectory of VGB before $\mathfrak{s}$ is visited $N+1$ times. The claim then follows from Lemma F.2. □

G Proof of Theorem 3.9

In this appendix we formally state the requisite definitions for Theorem 3.9, and prove the theorem. In Appendix G.1, we formalize the definition of a myopic multi-particle method. In Appendix G.2 we state Theorem G.5, which directly implies Theorem 3.9. In Appendix G.3, we provide intuition for and formally prove Theorem G.5.

G.1 Myopic Multi-Particle Methods

Definition G.1 (Myopic multi-particle method)**.** *A myopic multi-particle method with N particles is defined as follows. The algorithm computes a stochastic sequence of multisets $\mathcal{S}_0, \dots, \mathcal{S}_H$ where $\mathcal{S}_h \subset \mathcal{A}^h$ for all $0 \le h \le H$, satisfying the following properties:*

1. *It holds almost surely that for all $1 \le h \le H$, for each $y_{1:h} \in \mathcal{S}_h$, we have $y_{1:h-1} \in \mathcal{S}_h$.*
2. *$|\mathcal{S}_h| \le N$ for all h almost surely.*
3. *The law of $(\mathcal{S}_0, \dots, \mathcal{S}_h)$ is determined by $(\pi_{\mathsf{ref}}(a_{1:h}) : a_{1:h} \in \mathcal{A}^h)$ and $(\widehat{V}(a_{1:k}) : a_{1:k} \in \mathcal{A}^k, k \le h)$ for all h.*

The algorithm then (stochastically) outputs any element of $\mathcal{S}_H$.

The first condition is a natural formalization of what it means to be a multi-particle method, i.e. a particle in $\mathcal{S}_h$ cannot appear "out of nowhere"; rather, it must have a parent in $\mathcal{S}_{h-1}$. The second condition asserts that the number of particles is always at most N. The third condition is the "myopic" condition, which enforces that, at the level of distributions, the algorithm does not use information from later steps $k > h$ to determine the multiset $\mathcal{S}_h$.

Remark G.2 (Sequential Monte Carlo and variants)**.** *It is straightforward from Algorithm 1 to see that Sequential Monte Carlo is indeed a myopic multi-particle method, as is Sequential Monte Carlo with Rejection Sampling (Algorithm 2).*

Remark G.3 (Computational efficiency)**.** *We remark that Definition G.1 does not enforce any bound on computational efficiency or query complexity; in particular, a myopic multi-particle method is allowed to examine $\widehat{V}(a_{1:h})$ for all $a_{1:h} \in \mathcal{A}^h$ before deciding on the set of particles $\mathcal{S}_h$; it is not restricted to e.g. only querying the children of particles in $\mathcal{S}_{h-1}$. This only makes our impossibility result stronger. That said, it is an interesting open question to prove an analogous result to Theorem 3.9 that applies to* all *algorithms that make $o(H^2)$ queries to π_{ref} and $\widehat{V}$. Such a result would be formally incomparable to Theorem 3.9.*

G.2 Specialized Setting and Formal Theorem Statement

Since we are proving a lower bound, it suffices to consider a special case of the general setting from Section 2. Thus, we restrict to the autoregressive generation setting (Remark 2.1) and moreover set $\pi_{\mathsf{ref}} := \mathsf{Unif}(\mathcal{A}^H)$ where $\mathcal{A} = \{0, 1\}$.

We define a problem instance by specifying some distribution $\mu \in \Delta(\mathcal{A}^H)$ and functions $(\mu_h)_h$ with $\mu_h : \mathcal{A}^h \to \mathbb{R}_{\geq 0}$. These induce $V^\star, \widehat{V}$ by

$$V^\star(a_{1:h}) := \frac{\mu(a_{1:h})}{\pi_{\mathsf{ref}}(a_{1:h})}$$

and

$$\widehat{V}(a_{1:h}) := \frac{\widehat{\mu}_h(a_{1:h})}{\pi_{\mathsf{ref}}(a_{1:h})}.$$

Note that the goal distribution $\pi^\star$ is precisely μ, and the third condition of Definition G.1 precisely enforces that the law of $(\mathcal{S}_0, \dots, \mathcal{S}_h)$ is determined by $(\widehat{\mu}_0, \dots, \widehat{\mu}_h)$. For Assumption 3.1 to hold with parameter C_{act}, it suffices to have $\mu(a_{h+1} \mid a_{1:h}) \leq C_{\mathsf{act}}/2$ for all $a_{1:h+1} \in \mathcal{A}^{h+1}$. Moreover, Assumption 3.3 holds with parameter C_∞ if and only if $\widehat{\mu}$ is C_∞-accurate for μ in the following sense:

Definition G.4. *We say that $\widehat{\mu}$ is κ-accurate for $\mu \in \Delta(\mathcal{A}^H)$ if it holds that*

$$\frac{\widehat{\mu}(a_{1:h})}{\mu(a_{1:h})} \in [1/\kappa, \kappa]$$

for all $a_{1:h} \in \mathcal{A}^h$, and moreover $\widehat{\mu}(a_{1:H}) = \mu(a_{1:H})$ for all $a_{1:H} \in \mathcal{A}^H$.

We can now formally state the main result of this appendix, which directly implies Theorem 3.9.

Theorem G.5 (Lower bound for myopic multi-particle methods (full version))**.** *Define $N(H) := \log(H)/(8 \log\log(H))$ for all $H \in \mathbb{N}$. There is no myopic multi-particle method $\mathtt{Alg}$ with the following guarantee: fix H sufficiently large. $\mathtt{Alg}$ uses at most $N(H)$ particles. For any $\mu \in \Delta(\mathcal{A}^H)$, for any e^3-accurate $\widehat{\mu} : \mathcal{A}^{\leq H} \to \mathbb{R}_{\geq 0}$, the execution of $\mathtt{Alg}$ on $\widehat{\mu}$ has output $o_{1:H}$ satisfying*

$$\mathbb{P}_{o_{1:H} \sim \mathtt{Alg}}[o_{1:H} \in \mathcal{E}] \geq H^{-1}$$

for all events $\mathcal{E} \subset \mathcal{A}^H$ with $\mu(\mathcal{E}) \geq 1/2$.

Proof of Theorem 3.9. Immediate from Theorem G.5 and the preceding discussion on Assumptions 3.1 and 3.3. □

G.3 Proof of Theorem G.5

G.3.1 Intuition and Formal Construction

To prove Theorem G.5, we construct a family of problem instances (each specified by a choice of μ and $\widehat{\mu}_0, \dots, \widehat{\mu}_H$) such that any myopic multi-particle method with at most N particles must fail on at least one instance in the family. We do so recursively, by first constructing an instance for $N = 1$, then $N = 2$, and so forth. The analysis will be similarly inductive.

Let $\mathfrak{H} : \mathbb{N} \to \mathbb{N}$ be a function to be determined. Assume that $\mathfrak{H}(N)$ is a multiple of $\mathfrak{H}(N-1) + 2$ for all $N > 1$. For each $N \in \mathbb{N}$ and $y^\star \in \mathcal{A}^{\mathfrak{H}(N)}$, we define $\mu^N(\cdot; y^\star) \in \Delta(\mathcal{A}^{\mathfrak{H}(N)})$ and $\widehat{\mu}^N(\cdot; y^\star) : \mathcal{A}^{\leq \mathfrak{H}(N)} \to \mathbb{R}_{\geq 0}$ recursively (in N) as follows.

Intuition for construction. First, we provide intuition for the $N = 1$ construction, which essentially follows Example 3.2 in Rohatgi et al. [43]. We will have $\mu^1(\cdot; y^\star)$ be a product distribution where the marginal $\mu^1(y_h)$ is biased towards $y^\star_h$. We will then have $\widehat{\mu}^1(\cdot; y^\star)$ be "delayed" compared to $\mu^1(\cdot; y^\star)$, i.e. the marginal at step h is biased towards $y^\star_{h-1}$ instead of $y^\star_h$. Thus, the data of $\widehat{\mu}^1_1(\cdot; y^\star), \dots, \widehat{\mu}^1_h(\cdot; y^\star)$ is independent of $y^\star_h$, and so as formalized in Lemma G.9, any myopic multi-particle method with 1 particle must incur large sampling error.

Next, we provide intuition for the $N > 1$ construction. The horizon $H = \mathfrak{H}(N)$ is divided into blocks of size $K = \mathfrak{H}(N-1) + 2$. Again $\mu^N(\cdot; y^\star)$ is a product distribution where the marginal at step h is biased towards $y^\star_h$. In block $[iK+1, (i+1)K]$, we define $\widehat{\mu}^N(\cdot; y^\star)$ by "delaying" the dependence on $y^\star_{iK+1}$ to step $(i+1)K$, and embedding a copy of the $(N-1)$-particle construction in the intervening steps. Intuitively, any myopic multi-particle method with N particles has two choices at step $iK+1$: either irrevocably guess $y^\star_{iK+1}$ (i.e. all particles $y \in \mathcal{S}_{iK+1}$ will have the

same value of y_{iK+1}), or split up the particles so that some have different values of y_{iK+1}. The basic idea, formalized in Lemma G.11, is that both choices lead to sampling error: the former because no data about $y^\star_{iK+1}$ is available at step $iK+1$, and the latter by the inductive hypothesis on the $(N-1)$-particle construction. The multiple blocks are needed to amplify the error induced by the first choice.

Making this argument precise requires a careful probabilistic analysis to ensure that the errors in the two cases cannot "cancel out". Before proceeding to this analysis, we make the construction formal.

Let $\gamma_1, \gamma_2, \cdots \in (0, 1/2)$ be a sequence of parameters to be determined later.

Formal construction: base case ($N = 1$). Write $H = \mathfrak{H}(1)$. We define

$$\mu^1(y_h \mid y_{1:h-1}; y^\star) := (1/2 - \gamma_1)^{\mathbb{I}[y_h \neq y^\star_h]}(1/2 + \gamma_1)^{\mathbb{I}[y_h = y^\star_h]}$$

and

$$\widehat{\mu}^1(y_h \mid y_{1:h-1}; y^\star) := \begin{cases} 1/2 & \text{if } h = 1 \\ \mu_1(y_{h-1} \mid y_{1:h-2}; y^\star) & \text{if } 1 < h < H \\ 2\mu_1(y_{H-1:H} \mid y_{1:H-2}; y^\star) & \text{if } h = \mathfrak{H}(1) \end{cases}.$$

Formal construction: recursive case ($N > 1$). Define $K := \mathfrak{H}(N-1) + 2$. For any nonnegative integer i and $h \in [\mathfrak{H}(N)]$ with $iK < h \leq (i+1)K$, we define

$$\mu^N(y_h \mid y_{1:h-1}; y^\star) := \begin{cases} (1/2 - \gamma_N)^{\mathbb{I}[y_h \neq y^\star_h]}(1/2 + \gamma_N)^{\mathbb{I}[y_h = y^\star_h]} & \text{if } h = iK+1 \text{ or } h = (i+1)K \\ \mu^{N-1}(y_h \mid y_{iK+2:h-1}; y^\star_{iK+2:(i+1)K-1}) & \text{otherwise} \end{cases}$$

and

$$\widehat{\mu}^N(y_h \mid y_{1:h-1}; y^\star) := \begin{cases} 1/2 & \text{if } h = iK+1 \\ \widehat{\mu}^{N-1}(y_h \mid y_{iK+2:h-1}; y^\star_{iK+2:(i+1)K-1}) & \text{if } iK+1 < h < (i+1)K \\ 2\mu^N(y_{iK+1} \mid y_{1:iK}; y^\star)\mu^N(y_{(i+1)K} \mid y_{1:(i+1)K-1}; y^\star) & \text{if } h = (i+1)K \end{cases}$$

In the following lemma we check that the desired accuracy condition is satisfied for all N and $y^\star$.

Lemma G.6. *For any $N \geq 1$ and $y^\star \in \mathcal{A}^{\mathfrak{H}(N)}$, it holds that $\widehat{\mu}^N(\cdot; y^\star)$ is $\exp(3\sum_{i=1}^N \gamma_i)$-accurate for $\mu^N(\cdot; y^\star)$.*

Proof. We proceed by induction on N. Fix $N = 1$. Set $H = \mathfrak{H}(1)$. We observe that $\widehat{\mu}^1(y_{1:H}; y^\star) = \mu^1(y_{1:H}; y^\star)$ and, for any $h < H$,

$$\begin{aligned} \frac{\widehat{\mu}^1(y_{1:h}; y^\star)}{\mu_1(y_{1:h}; y^\star)} &= \frac{1}{2\mu^1(y_h \mid y_{1:h-1}; y^\star)} \\ &\in [1/(1+2\gamma_1), 1/(1-2\gamma_1)] \\ &\in [e^{-3\gamma_1}, e^{3\gamma_1}]. \end{aligned}$$

Fix $N > 1$ and suppose that the claim holds for $\widehat{\mu}^{N-1}$. Set $H = \mathfrak{H}(N)$ and $K = \mathfrak{H}(N-1) + 2$. For any $h \in [H]$, if $h = (i+1)K$ for some i, then

$$\begin{aligned} &\widehat{\mu}^N(y_{iK+1:(i+1)K} \mid y_{1:iK}; y^\star) \\ &= \frac{1}{2} \cdot \left(\prod_{h'=iK+2}^{(i+1)K-1} \widehat{\mu}^{N-1}(y_{h'} \mid y_{iK+2:h'-1}; y^\star_{iK+2:(i+1)K-1}) \right) \\ &\quad \cdot 2\mu^N(y_{iK+1} \mid y_{1:iK}; y^\star)\mu^N(y_{(i+1)K} \mid y_{1:(i+1)K-1}; y^\star) \\ &= \mu^{N-1}(y_{iK+2:(i+1)K-1}; y^\star_{iK+2:(i+1)K-1}) \cdot \mu^N(y_{iK+1} \mid y_{1:iK}; y^\star)\mu^N(y_{(i+1)K} \mid y_{1:(i+1)K-1}; y^\star) \\ &= \mu^N(y_{iK+2:(i+1)K-1} \mid y_{1:iK+1}; y^\star) \cdot \mu^N(y_{iK+1} \mid y_{1:iK}; y^\star)\mu^N(y_{(i+1)K} \mid y_{1:(i+1)K-1}; y^\star) \\ &= \mu^N(y_{iK+1:(i+1)K} \mid y_{1:iK}; y^\star). \end{aligned}$$

Thus, it holds that $\widehat{\mu}^N(y_{1:(i+1)K}; y^\star) = \mu^N(y_{1:(i+1)K}; y^\star)$, and in particular, $\widehat{\mu}^N(y_{1:H}; y^\star) = \mu^N(y_{1:H}; y^\star)$. For any $iK + 1 \le h < (i+1)K$, we have

$$\begin{aligned}
\widehat{\mu}^N(y_{1:h}; y^\star) &= \mu^N(y_{1:iK}; y^\star)\widehat{\mu}^N(y_{iK+1:h} \mid y_{1:iK}; y^\star) \\
&= \mu^N(y_{1:iK}; y^\star) \cdot \frac{1}{2} \cdot \widehat{\mu}^{N-1}(y_{iK+2:h}; y^\star_{iK+2:(i+1)K-1}) \\
&\in [e^{-3\sum_{n=1}^{N-1}\gamma_n}, e^{3\sum_{n=1}^{N-1}\gamma_n}] \cdot \mu^N(y_{1:iK}; y^\star) \cdot \frac{1}{2} \cdot \mu^{N-1}(y_{iK+2:h}; y^\star_{iK+2:(i+1)K-1}) \\
&\in [e^{-3\sum_{n=1}^{N}\gamma_n}, e^{3\sum_{n=1}^{N}\gamma_n}] \cdot \mu^N(y_{1:iK}; y^\star) \cdot \mu^N(y_{iK+1} \mid y_{1:iK}; y^\star) \cdot \mu^{N-1}(y_{iK+2:h}; y^\star_{iK+2:(i+1)K-1}) \\
&\in [e^{-3\sum_{n=1}^{N}\gamma_n}, e^{3\sum_{n=1}^{N}\gamma_n}] \cdot \mu^N(y_{1:h}; y^\star)
\end{aligned}$$

which completes the proof. □

To show that any myopic multi-particle method with N particles fails to cover $\mu^N(\cdot; y^\star)$ (for some choice of $y^\star$), we will show that there are events $\mathcal{F}_N(y^\star)$ that are extremely likely under $\mu^N(\cdot; y^\star)$ (for any $y^\star$) but extremely unlikely under the execution of any myopic multi-particle method with N particles (for a uniformly random choice of $y^\star$). Definition G.7 formally defines these events. Definition G.8 formalizes the claim about these events that we will prove inductively.

Definition G.7 (Key events for analysis)**.** *Fix $N = 1$ and $H = \mathfrak{H}(1)$. We define*

$$\mathcal{F}_1(y^\star) := \{o_{1:H} : \#\{1 < h < H : o_h = y^\star_h\} \ge (1/2 + \gamma_1/2)(H-2)\}.$$

Fix $N > 1$ and write $H := \mathfrak{H}(N)$ and $K := \mathfrak{H}(N-1) + 2$. We define sets $\mathcal{F}_N(y^\star)$ as follows:

$$\mathcal{F}^\circ_N(y^\star) := \bigcap_i \left\{ o_{1:H} : o_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \right\}$$

$$\mathcal{F}^\square_N(y^\star) := \{o_{1:H} : \#\{i : o_{iK+1} = y^\star_{iK+1}\} \ge 1/2 + \gamma_N/2\}$$

$$\mathcal{F}_N(y^\star) := \mathcal{F}^\circ_N(y^\star) \cap \mathcal{F}^\square_N(y^\star).$$

Definition G.8 (Inductive hypothesis for analysis)**.** *Fix $N \ge 1$. For parameter $\delta_{N,1}, \delta_{N,2} \in (0,1)$, we let $\mathsf{IH}_N(\delta_{N,1}, \delta_{N,2})$ denote the following hypothesis. There are sets $\mathcal{F}_N(y^\star) \subseteq \mathcal{A}^{\mathfrak{H}(N)}$ with the following properties:*

- *For all $y^\star \in \mathcal{A}^{\mathfrak{H}(N)}$, it holds that*

$$\mu^N(\mathcal{F}_N(y^\star); y^\star) \ge 1 - \delta_{N,1}$$

- *For any multi-particle method* `Alg` *with N particles, it holds that*

$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(N)})} \mathbb{P}_{o \sim \mathtt{Alg}}[o \in \mathcal{F}_N(y^\star)] \le \delta_{N,2}$$

  where the probability is over outputs o of `Alg` *executed on instance $\widehat{\mu}^N(\cdot; y^\star)$.*

G.3.2 Base Case for Analysis

Lemma G.9. *For any $y^\star \in \mathcal{A}^{\mathfrak{H}(1)}$, it holds that*

$$\mu^1(\mathcal{F}_1(y^\star); y^\star) \ge 1 - \exp\left(-\gamma_1^2 \frac{\mathfrak{H}(1) - 2}{16}\right).$$

Moreover, for any multi-particle method `Alg` *with 1 particle,*

$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(1)})} \mathbb{P}_{o \sim \mathtt{Alg}}[o \in \mathcal{F}_1(y^\star)] \le \exp\left(-\gamma_1^2 \frac{\mathfrak{H}(1) - 2}{12}\right)$$

where `Alg` *is executed on instance $\widehat{\mu}^1(\cdot; y^\star)$.*

Proof. Let $X \sim \mu^1(\cdot; y^\star)$. Then $X_1, \dots, X_H$ are independent with $\mathbb{P}[X_h = y^\star_h] = 1/2 + \gamma_1$ for each $h \in [H]$. It follows from a Chernoff bound that

$$\begin{aligned}\mu^1(\mathcal{F}_1(y^\star); y^\star) &= \mathbb{P}[\#\{1 < h < H : X_h = y^\star_h\} > (1/2 + \gamma_1/2)H] \\ &\geq 1 - \exp(-\gamma_1^2(\mathfrak{H}(1) - 2)/16)\end{aligned}$$

which proves the first claim of the lemma. Next, fix any multi-particle method $\texttt{Alg}$ with 1 particle. Sample $y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(1)})$ and let $\mathcal{S}_0, \dots, \mathcal{S}_H$ be the stochastic sequence of multisets obtained by executing $\texttt{Alg}$ on $\widehat{\mu}^1(\cdot; y^\star)$. Let o be the output of $\texttt{Alg}$. Then $\mathcal{S}_h = \{o_{1:h}\}$ for all $0 \leq h \leq H$. Since the law of $\mathcal{S}_{0:h}$ is determined by $(\widehat{\mu}^1_0(\cdot; y^\star), \dots, \widehat{\mu}^1_h(\cdot; y^\star))$, it holds from the definition of $\widehat{\mu}^1(\cdot; y^\star)$ that $(\mathcal{S}_{0:h}, y^\star_{1:h-1})$ is independent of $y^\star_h$ for each $1 < h < H$. Thus, $y^\star_h$ is independent of $\mathcal{S}_h$ conditioned on $(\mathcal{S}_{0:h-1}, y^\star_{1:h-1})$, and so we have

$$\mathbb{P}[o_h = y^\star_h \mid \mathcal{S}_{0:h-1}, y^\star_{1:h-1}] = \frac{1}{2}$$

for all $1 < h < H$. The second claim of the lemma then follows from Azuma's inequality. □

G.3.3 Inductive Step for Analysis

Lemma G.10. *Fix $N > 1$. Assume that $\mathsf{IH}_{N-1}(\delta_{N-1,1}, \delta_{N-1,2})$ holds. Suppose $\mathfrak{H}(N-1) \geq 2$. For any $y^\star \in \mathcal{A}^{\mathfrak{H}(N)}$, it holds that*

$$\mu^N(\mathcal{F}_N(y^\star); y^\star) \geq 1 - \frac{\mathfrak{H}(N)}{\mathfrak{H}(N-1)}\delta_{N-1,1} - \exp\left(-\gamma_N^2 \frac{\mathfrak{H}(N)}{16 \cdot \mathfrak{H}(N-1)}\right)$$

Proof. Set $H := \mathfrak{H}(N)$ and $K := \mathfrak{H}(N-1) + 2$. Let $o \sim \mu^N(\cdot; y^\star)$. Then for each i, the law of $o_{iK+2:(i+1)K-1}$ is precisely $\mu^{N-1}(\cdot; y^\star_{iK+2:(i+1)K-1})$. Thus,

$$\begin{aligned}&\mu^N(\mathcal{F}^\circ_N(y^\star); y^\star) \\ &\geq 1 - \sum_i \mu^N(o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})) \\ &= 1 - \frac{H}{K}\Big(1 - \mu^{N-1}(\mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}); y^\star_{iK+2:(i+1)K-1})\Big) \\ &\geq 1 - \frac{H}{K}\delta_{N-1,1} \\ &\geq 1 - \frac{\mathfrak{H}(N)}{\mathfrak{H}(N-1)}\delta_{N-1,1}.\end{aligned}$$

Next, note that the random variables $(o_{iK+1})_i$ are independent and $\mathbb{P}[o_{iK+1} = y^\star_{iK+1}] = 1/2 + \gamma_N$ for each i. Thus,

$$\mu^N(\mathcal{F}^\square_N(y^\star); y^\star) \geq 1 - \exp(-\gamma_N^2(H/K)/8) \geq 1 - \exp(-\gamma_N^2(\mathfrak{H}(N)/\mathfrak{H}(N-1))/16$$

by Hoeffding's inequality, so long as $\mathfrak{H}(N-1) \geq 2$. The lemma follows from the union bound. □

Lemma G.11. *Fix $N > 1$ and $\delta_{N-1,1}, \delta_{N-1,2}, \delta_{N,2} \in (0, 1/2)$. Assume that $\mathsf{IH}_{N-1}(\delta_{N-1,1}, \delta_{N-1,2})$ holds. Set $H := \mathfrak{H}(N)$ and $K := \mathfrak{H}(N-1) + 2$. If*

$$\gamma_N > 4\delta_{N-1,2} + \sqrt{\frac{8K\log(3/\delta_{N,2})}{H}} + \frac{8K}{H}\log(6/\delta_{N,2}),$$

then for any multi-particle method $\texttt{Alg}$ with N particles, it holds that

$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(N)})} \mathbb{P}_{o \sim \texttt{Alg}}[o \in \mathcal{F}_N(y^\star)] \leq \delta_{N,2}$$

where $\texttt{Alg}$ is executed on instance $\widehat{\mu}^N(\cdot; y^\star)$.

Proof. Consider the random process induced by sampling $y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{h}(N)})$ and executing $\mathtt{Alg}$ on $\widehat{\mu}^N(\cdot; y^\star)$. Fix i and condition on $\mathcal{S}_{0:iK}$. Let $\mathcal{E}^i$ be the event that $y_{iK+1} \neq y'_{iK+1}$ for some $y, y' \in \mathcal{S}_{(i+1)K-1}$. For each $b \in \{0,1\}$, define process $\mathcal{T}^{(0,b)}, \ldots, \mathcal{T}^{(K-2,b)}$ where $\mathcal{T}^{(0,b)} = \{\varnothing\}$ and, for each $h' > 0$, $\mathcal{T}^{(h',b)}$ consists of the first $\leq N-1$ elements of

$$\{y_{iK+2:h'+iK+1} : y_{1:h'+iK+1} \in \mathcal{S}_{h'+iK+1} \wedge y_{iK+1} = b \wedge y_{iK+2:h'+iK} \in \mathcal{T}^{(h'-1,b)}\}$$

according to some arbitrary ordering of $\mathcal{A}^{h'}$. Observe that $\mathcal{T}^{(0,b)}, \ldots, \mathcal{T}^{(K-2,b)}$ describes the execution of a myopic multi-particle method with $(N-1)$ particles on $\widehat{\mu}^{N-1}(\cdot; y^\star_{iK+2:(i+1)K-1})$. To see why, we check the conditions of Definition G.1. First, it holds almost surely that $|\mathcal{T}^{(h',b)}| \leq N-1$ for all h'. Second, by construction, if $y_{1:h'} \in \mathcal{T}^{(h',b)}$ then $y_{1:h'-1} \in \mathcal{T}^{(h'-1,b)}$. Third, for any $0 \leq h \leq K-2$, the law of $\mathcal{T}^{(0,b)}, \ldots, \mathcal{T}^{(h',b)}$ is determined by the law of $\mathcal{S}_{0:h'+iK+1}$, which by definition of $\mathtt{Alg}$ is determined by $(\widehat{\mu}^N_h(\cdot; y^\star) : h \leq h' + iK + 1)\}$, which by definition of $\widehat{\mu}^N$ is determined by $(\widehat{\mu}^{N-1}_k(\cdot; y^\star_{iK+2:(i+1)K-1}) : 0 \leq k \leq h')$ (together with randomness independent of $y^\star_{iK+2:(i+1)K-1}$).

It follows from the inductive hypothesis that

$$\mathbb{P}[\exists y \in \mathcal{T}^{(K-2,b)} \cap \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})] \leq \delta_{N-1,2}.$$

In the event $\mathcal{E}^i$, it holds almost surely that

$$\{y_{iK+2:(i+1)K-1} : y_{1:(i+1)K-1} \in \mathcal{S}_{(i+1)K-1}\} = \mathcal{T}^{(K-2,0)} \cup \mathcal{T}^{(K-2,1)}.$$

Moreover, the output $o_{1:H}$ of $\mathtt{Alg}$ satisfies $o_{1:(i+1)K-1} \in \mathcal{S}_{(i+1)K-1}$ almost surely, so

$$\begin{aligned}
&\mathbb{P}[o_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\
&\leq \mathbb{P}[\exists y_{1:(i+1)K-1} \in \mathcal{S}_{(i+1)K-1} : y_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\
&\leq \mathbb{P}[\overline{\mathcal{E}}^i \mid \mathcal{S}_{0:iK}] + \mathbb{P}[\exists y \in \mathcal{T}^{(K-2,0)} \cup \mathcal{T}^{(K-2,1)} : y_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\
&\leq \mathbb{P}[\overline{\mathcal{E}}^i \mid \mathcal{S}_{0:iK}] + 2\delta_{N-1,2}.
\end{aligned}$$

Next, define $v_{iK+1} = \mathrm{mode}\{y_{iK+1} : y \in \mathcal{S}_{(i+1)K-1}\}$. Observe that in the event $\overline{\mathcal{E}}^i$, it holds that $o_{iK+1} = v_{iK+1}$. Observe that $\mathcal{S}_{(i+1)K-1}$ is independent of $y^\star_{iK+1}$, so that

$$\mathbb{P}[v_{iK+1} = y^\star_{iK+1} \mid \mathcal{S}_{0:iK}] = \frac{1}{2}.$$

Define random variables

$$W_i := \mathbb{I}[v_{iK+1} = y^\star_{iK+1}] - \mathbb{P}[v_{iK+1} = y^\star_{iK+1} \mid \mathcal{S}_{0:iK}].$$

Then $W_0, \ldots, W_{H/K}$ is a martingale difference sequence under filtration induced by $(\mathcal{S}_{0:iK})_{i=0}^{H/K-1}$, so for any $\delta \in (0,1)$,

$$\mathbb{P}\left[\sum_{i=0}^{H/K-1} W_i > \sqrt{8(H/K)\log(1/\delta)}\right] \leq \delta.$$

Thus, with probability at least $1-\delta$, it holds that

$$\begin{aligned}
&\frac{K}{H}\sum_{i=0}^{H/K-1} \mathbb{I}[o_{iK+1} = y^\star_{iK+1}] \\
&\leq \frac{K}{H}\sum_{i=0}^{H/K-1} \mathbb{I}[v_{iK+1} = y^\star_{iK+1}] + \frac{K}{H}\sum_{i=0}^{H/K-1} \mathbb{I}[o_{iK+1} \neq v_{iK+1}] \\
&\leq \frac{1}{2} + \sqrt{\frac{8K\log(1/\delta)}{H}} + \frac{K}{H}\sum_{i=0}^{H/K-1} \mathbb{I}[\mathcal{E}^i].
\end{aligned}$$

Define random variables

$$Y_i := \mathbb{I}[\mathcal{E}^i] - \mathbb{P}[\mathcal{E}^i \mid \mathcal{S}_{0:iK}].$$

Then $Y_0, \dots, Y_{H/K-1}$ is a martingale difference sequence under filtration induced by $(\mathcal{S}_{0:iK})_{i=0}^{H/K-1}$. Thus, with probability at least $1-\delta$, it holds that

$$\begin{aligned}\frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[\mathcal{E}^i] &\le \sqrt{\frac{8K\log(1/\delta)}{H}} + \frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{P}[\mathcal{E}^i \mid \mathcal{S}_{0:iK}] \\ &\le \sqrt{\frac{8K\log(1/\delta)}{H}} + 2\delta_{N-1,2} + \frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{P}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}].\end{aligned}$$

By Freedman's inequality, it holds with probability at least $1-\delta$ that

$$\begin{aligned}&\sum_{i=0}^{H/K-1}\mathbb{P}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\ &\le 2\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})] + 8\log(2/\delta).\end{aligned}$$

Combining the above inequalities with a union bound, we conclude that with probability at least $1-3\delta$,

$$\begin{aligned}&\frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+1} = y^\star_{iK+1}] \\ &\le \frac{1}{2} + 2\delta_{N-1,2} + \sqrt{\frac{32K\log(1/\delta)}{H}} + \frac{8K}{H}\log(2/\delta) \\ &\quad + \frac{2K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})].\end{aligned}$$

Condition on this event. If we moreover have $o_{1:H} \in \mathcal{F}_N(y^\star)$, then we have both

$$\frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+1} = y^\star_{iK+1}] \ge \frac{1}{2} + \frac{\gamma_N}{2}$$

and

$$\frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})] = 0.$$

But then

$$\frac{\gamma_N}{2} \le 2\delta_{N-1,2} + \sqrt{\frac{8K\log(1/\delta)}{H}} + \frac{8K}{H}\log(2/\delta).$$

Set $\delta := \delta_{N,2}/3$. Then this bound contradicts the lemma assumption, so in fact $o_{1:H} \notin \mathcal{F}_N(y^\star)$ almost surely (under the preceding condition). It follows that $\mathbb{P}[o_{1:H} \in \mathcal{F}_N(y^\star)] \le 3\delta = \delta_{N,2}$ as claimed. □

G.3.4 Completing the Induction

Fix some $N \ge 2$. Define $\mathfrak{H}(1) := N^8$ and $\mathfrak{H}(n) := N^8(\mathfrak{H}(n-1)+2)$ for all $n > 1$, so that $\mathfrak{H}(n)$ is a multiple of $\mathfrak{H}(n-1)+2$ as required by the construction, and $N^{8n} \le \mathfrak{H}(n) \le 2N^{8n}$ for all n. Define $\gamma_n := 1/N$ for all n.

Lemma G.12. *It holds that*

$$\mu^N(\mathcal{F}_N(y^\star); y^\star) \ge 1 - \exp(-\Omega(N^2))$$

for all $y^\star$, and

$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(N)})}\mathbb{P}_{o\sim\mathtt{Alg}}[o \in \mathcal{F}_N(y^\star)] \le 4\exp(-N^2).$$

Proof. We have from Lemmas G.9 and G.10 that

$$\mu^N(\mathcal{F}_N(y^\star); y^\star) \geq 1 - \sum_{n=1}^{N} \frac{\mathfrak{H}(N)}{\mathfrak{H}(n-1)} \exp(-\gamma_n^2 \cdot \Omega(\mathfrak{H}(n)/\mathfrak{H}(n-1))) \geq 1 - 2N^{8N+1} \exp(-\Omega(N^2)).$$

Next, observe that if we set $\delta_{n,2} := 4\exp(-N^2)$ then the condition of Lemma G.11 is satisfied for each n, so long as N exceeds a sufficiently large constant. Thus, applying Lemma G.11 together with induction on N (with Lemma G.9 providing the guarantee for $N = 1$) proves that

$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(n)})} \mathbb{P}_{o \sim \mathtt{Alg}}[o \in \mathcal{F}_n(y^\star)] \leq 4\exp(-N^2)$$

for each $n \leq N$ as needed. □

Proof of Theorem G.5. Immediate from Lemmas G.6 and G.12, together with the fact that the N-particle construction has horizon $H = \mathfrak{H}(N) \in [N^{8N}, 2N^{8N}]$, and hence $N \geq \log(H)/(8\log\log(H))$. □

H Additional Experimental Details

In this section, we provide further details regarding our experimental setup.

H.1 Additional Evaluation Approaches

First, we give additional context and motivation for our choice of the evaluation metric of logprob-discrepancy introduced in Section 4.1.

One approach to evaluate SMC, which has been used in prior work, is to use the fact that the value $\widehat{W}_H$ computed in Algorithm 1 satisfies $\mathbb{E}[\widehat{W}_H] = Z$ and moreover the difference $\log Z - \mathbb{E}[\log \widehat{W}_H]$ is an upper bound on the KL divergence $D_{\mathsf{KL}}(\mathbb{E}[\widehat{\nu}_H] \,\|\, \pi^\star_H)$ between the output distribution of SMC, $\mathbb{E}[\widehat{\nu}_H]$, and the target distribution $\pi^\star_H$ (Proposition 5 of Zhao et al. [62]).[8] However, there are several pitfalls with this approach: first, this is only an *upper bound* and is not guaranteed to be tight; further, it only applies to SMC and so does not allow us to compare the performance of SMC to other baselines. Finally, even for the relatively simple setting in which we operate, the number N of particles is too small to get actual *distributional* closeness between $\mathbb{E}[\widehat{\nu}_H]$ and the target $\pi^\star_H$ (for instance, the action-level coverage is too large).

Thus, we aimed to use a weaker notion of distributional closeness that we can both (a) measure efficiently and (b) hope to have some bearing on the quality of samples produced by SMC. The $\mathcal{P}$-logprob discrepancy certainly satisfies (a). Moreover, since its error is one-sided in nature (i.e., $\mathrm{LPD}_{\mathcal{P}}(\nu, \nu) = 0$) and it takes on a positive value in all of our experiments, we believe it is a good candidate in the context of the prompt switching task.

H.2 PRM Accuracy Predicts SMC Performance

Below we discuss in more detail the experiment in Section 4.2 used to evaluate the relationship between PRM accuracy and SMC performance. To evaluate the influence of PRM accuracy on SMC performance, we fix the action-level coverage between $\pi_{\mathtt{ref}}$ and $\pi^\star$ (in particular, by fixing $\mathsf{p}_{\mathtt{ref}} = \mathsf{p}^\star$) and vary the PRM $\widehat{V}$. In particular, we consider a number $k = 50$ of instances of the prompt-switching task as described above: each instance $i \in [k]$ is specified by three prompts $((\mathsf{p}_{\mathtt{ref}})^{(i)}, (\mathsf{p}^\star)^{(i)}, \mathsf{p}^{(i)})$, as well as a scalar α used in Eq. (7) below. In this subsection, the reference $(\mathsf{p}_{\mathtt{ref}})^{(i)}$ and target $(\mathsf{p}^\star)^{(i)}$ prompts are shared across all instances, so we denote them by $\mathsf{p}_{\mathtt{ref}}, \mathsf{p}^\star$, respectively. Moreover, we in fact have $\mathsf{p}_{\mathtt{ref}} = \mathsf{p}^\star$—this prompt asks the model to write a short story (e.g., "Write a scene about a dragon negotiating peace with the last human kingdom.").[9] Each

[8] Further, in Lemma D.3 we show that the difference $|\widehat{W}_H - Z|$ upper bounds the *total variation distance* $D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H)$.

[9] While in this setting it is trivial to sample from $\mathsf{p}^\star$, we emphasize that our main goal here is studying the performance of SMC in terms of PRM accuracy as opposed to solving a challenging sampling problem.

instance $i \in [k]$ has a different value of $\widehat{V}$, denoted $\widehat{V}^{(i)}$, which is given as follows: we fix k prompts $\mathsf{p}^{(1)}, \ldots, \mathsf{p}^{(k)}$, which are each identical to $\mathsf{p}_{\mathtt{ref}}$ except they ask for the story in k different styles (e.g., "... Tell it as a news article."). For each such prompt $\mathsf{p}^{(i)}$, we let $\pi^{(i)}$ denote the distribution $\pi^{(i)}(a_{1:H}) = \mathsf{M}(a_{1:H} \mid \mathsf{p}^{(i)})$, which should be interpreted as a "perturbed" version of $\pi^\star$. Then we define

$$\widehat{V}^{(i)}(a_{1:h}) := \frac{\pi^\star(a_{1:h})}{\pi_{\mathtt{ref}}(a_{1:h})} \cdot \left(\frac{\pi^{(i)}(a_{1:h})}{\pi^\star(a_{1:h})}\right)^{(1-h/H)\cdot\alpha}, \tag{7}$$

for some parameter $\alpha > 0$ controlling the degree of accuracy of $\widehat{V}^{(i)}$. To get some intuition behind this definition, suppose $\alpha = 1$ for simplicity: for small h, $\widehat{V}^{(i)}(a_{1:h}) \approx \pi^{(i)}(a_{1:h})/\pi_{\mathtt{ref}}(a_{1:h})$, while at the final step $h = H$, $\widehat{V}^{(i)}(a_{1:H}) = V^\star(a_{1:H})$. In particular, at early steps, $\widehat{V}_h^{(i)}$ looks the value function of the "perturbed" target distribution $\pi^{(i)}$, meaning that for small h, $\widehat{V}_h^{(i)}$ is "misleading" in that it "points towards" $\pi^{(i)}$ instead of $\pi^\star$. For larger h, we attenuate the noise so as to ensure that sampling from $V_H^\star$ is information-theoretically possible (i.e., that $\widehat{V}_H^{(i)} = V_H^\star$).

Natural variation in the choice of the prompts $\mathsf{p}^{(i)}$ (which were generated by GPT-5) leads the resulting functions $\widehat{V}_h^{(i)}$ to have varying levels of accuracy with respect to $V_h^\star$. Define $\widehat{\pi}_h^{(i)} \in \Delta(\mathcal{A}^h)$ by $\widehat{\pi}_h^{(i)}(a_{1:h}) \propto \pi_{\mathtt{ref}}(a_{1:h}) \cdot \widehat{V}^{(i)}(a_{1:h})$. While estimating $D_{\chi^2}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$ is in principle possible, we found that the ratios $V^\star(a_{1:h})/\widehat{V}^{(i)}(a_{1:h})$ for the present prompt-switching task needed to estimate the χ^2 divergence are prohibitively high in variance. Thus, we instead estimate $D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$ to measure the divergence between $\pi_h^\star$ and $\widehat{\pi}_h^{(i)}$, which is much more stable because KL divergence can be written as an expectation of *log*-probabilities (see Appendix H.5 for further details).

In our experiments, we use Qwen3-0.6B as the language model M and set the horizon to $H = 64$ tokens. For each choice of $i \in [k]$ we run SMC with $N = 32$ particles given the PRM $\widehat{V}^{(i)}$, and sample from its output distribution $\widehat{\nu}_H^{(i)}$. In Figure 1a, we plot the values of $D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$ against the sampling error of SMC, measured via $\mathcal{P}$-logprob discrepancies $\mathrm{LPD}_{\mathcal{P}}(\mathbb{E}[\widehat{\nu}_H^{(i)}], \pi_H^\star)$ (estimated by running SMC for 200 trials). We observe a clear correlation between the two quantities, thus validating our theoretical findings that PRM accuracy (as measured by the divergence between $\widehat{\pi}_h$ and $\pi_h^\star$) predicts SMC's sampling error. In Appendix H.5 we repeat the experiment for additional choices of $\mathsf{p}_{\mathtt{ref}}, \mathsf{p}^\star, \mathsf{p}^{(i)}, \alpha$ and observe similar results.

H.3 Action-Level Coverage Predicts SMC Performance

Next we discuss in more detail the experiment in Section 4.2 used to evaluate the relationship between action-level coverage and SMC performance. To evaluate the influence of action-level coverage on SMC performance, we consider the same setup as in the previous subsection, except for each instance $i \in [k]$ we take the target distribution, denoted $(\pi^\star)^{(i)}$, to be the same as $\pi^{(i)}$. In particular, each instance is determined by the tuple $(\mathsf{p}_{\mathtt{ref}}, \mathsf{p}^{(i)}, \mathsf{p}^{(i)})$. Moreover, we take $\alpha = 0$, so that $\widehat{V}^{(i)} = (V^\star)^{(i)}$. In this setting, the *action-level coverage* depends on i. We again use Qwen3-0.6B, take the horizon to be $H = 32$ tokens, and run SMC with $N = 32$ particles.

While the action-level coverage between $(\pi^\star)^{(i)}$ and $\pi_{\mathtt{ref}}$ (in the L_∞ sense of Assumption 3.1) is extremely large and essentially impossible to compute, we note that $D_{\mathsf{KL}}\left((\pi^\star)^{(i)} \,\|\, \pi_{\mathtt{ref}}\right)$ can be viewed as an "average-case proxy" for action-level coverage in light of the chain rule, which gives

$$D_{\mathsf{KL}}\left((\pi^\star)^{(i)} \,\|\, \pi_{\mathtt{ref}}\right) = \sum_{h=1}^{H} \mathbb{E}_{a_{1:h} \sim (\pi^\star)_h^{(i)}}\left[\log \frac{(\pi^\star)_h^{(i)}(a_h \mid a_{1:h-1})}{\pi_{\mathtt{ref},h}(a_h \mid a_{1:h-1})}\right].$$

Figure 1b plots $\frac{1}{H} D_{\mathsf{KL}}\left((\pi^\star)^{(i)} \,\|\, \pi_{\mathtt{ref}}\right)$ against the $\mathcal{P}$-logprob discrepancies $\mathrm{LPD}_{\mathcal{P}}(\mathbb{E}[\widehat{\nu}_H^{(i)}], \pi_H^\star)$; we again observe a clear correlation between the two quantities, in line with our theory.

H.4 Scaling with the Number of Particles

In Figure 5, we show the performance of SMC as a function of the number of particles N for the prompt switching task, using a setup similar to that of Appendix H.2. In particular, we consider 50 different prompt tuples $(\mathsf{p}_{\mathsf{ref}}^{(i)}, (\mathsf{p}^\star)^{(i)}, \mathsf{p}^{(i)})$, and for each tuple we use a PRM $\widehat{V}^{(i)}$ defined from $\mathsf{p}^{(i)}$ as in Eq. (7) with $\alpha = 1$. In addition to SMC, we also plot the performance of (a) sequential importance sampling using $\widehat{V}^{(i)}$ (Appendix B), and (b) the "best-of-N" algorithm which samples N sequences from π_{ref} and outputs the one with the highest value according to the terminal reward $(V^\star)^{(i)}(a_{1:H})$. We observe that SMC consistently outperforms both baselines, and moreover that its performance improves as N increases; the latter conclusion is in line with our theoretical results.

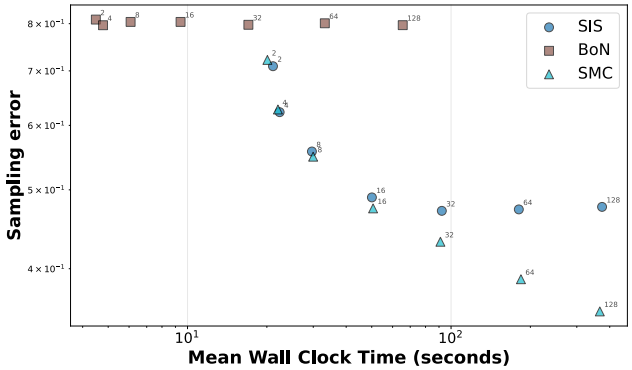


Figure 5: Performance of SMC on the prompt-switching task (Appendix H.4). Each point represents SMC with some number of particles $N \in \{2, 4, 8, \ldots, 128\}$. We average the sampling error and wall-clock time over all data points and all trials per data point. SMC consistently outperforms both sequential importance sampling (SIS) and Best-of-N (BoN) baselines.

H.5 Details for Appendix H.2

For the prompt switching task in Appendix H.2, the prompts $\mathsf{p}_{\mathsf{ref}}, \mathsf{p}^{(i)}$ were all generated with GPT-5.2 (in particular, the gpt-5.2_2025-12-11 release).

To estimate $D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$, we write

$$\begin{aligned} D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right) &= \mathbb{E}_{a_{1:h}\sim\pi_h^\star}\left[\log \frac{\pi_h^\star(a_{1:h})}{\widehat{\pi}_h^{(i)}(a_{1:h})}\right] \\ &= \mathbb{E}_{a_{1:h}\sim\pi_h^\star}\left[(1-h/H)\cdot\alpha\cdot\log\frac{\pi^\star(a_{1:h})}{\pi^{(i)}(a_{1:h})}\right] + \log \widehat{Z}_h^{(i)}, \end{aligned} \tag{8}$$

where

$$\widehat{Z}_h^{(i)} := \mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}^{(i)}(a_{1:h})] = \mathbb{E}_{\pi^\star}\left[\left(\frac{\pi_h^{(i)}(a_{1:h})}{\pi_h^\star(a_{1:h})}\right)^{(1-h/H)\cdot\alpha}\right]. \tag{9}$$

We estimate the first term in Eq. (8) via Monte Carlo rollouts from $\pi_h^\star$, and estimate $\log \widehat{Z}_h^{(i)}$ via Monte Carlo rollouts from $\pi^\star$, per Eq. (9) (each with 1000 samples).

Notice that our estimate of $\widehat{Z}_h^{(i)}$ is likely very noisy, since there is no logarithm inside the expectation. When we repeated the experiments without adding (our estimate of) the term $\log \widehat{Z}_h^{(i)}$ at all in Eq. (8), we observed similar results. Thus, the noise in our estimate of $\widehat{Z}_h^{(i)}$ seems unlikely to have affected our conclusions.

Details for the figures. In Figure 1a, we use the PRM defined in Eq. (7) with $\alpha = 2$. Moreover, the KL divergences are measured at step $h = 32$ (recall that $H = 64$); we observed similar results with

other values of h. The experiments were run on 32GB NVIDIA V100 GPUs. The Pearson correlation coefficient for the points plotted in Figure 1a is 0.81.

In Figure 8, we plot additional runs of the same experiment as in Figure 1a, with 3 additional different choices of the dataset (i.e., different choices of $\mathsf{p}_{\mathtt{ref}}$ and $\mathsf{p}^{(i)}$). Finally, in Figure 9, we repeat the same experiment for all 4 datasets but with $\alpha = 1$.

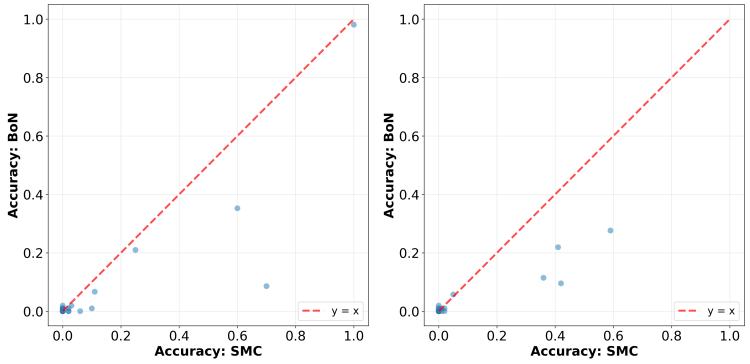
H.6 Details for Appendix H.3

In Figure 7, we repeat the experiment in Figure 1b with 4 additional choices for the dataset (i.e., choices of $\mathsf{p}_{\mathtt{ref}}$ and $\mathsf{p}^{(i)}$, $i \in [k]$). The Pearson correlation coefficients r are shown in the subcaptions of each subfigure. The experiments were run on 32GB NVIDIA V100 GPUs.

H.7 SMC vs. Best-of-N with Qwen Math PRM

In our experiments on math problems in Section 4.3, we used Qwen2.5-1B-Instruct as the base model $\pi_{\mathtt{ref}}$ and Qwen2.5-Math-PRM-7B as the PRM $\widehat{V}$. For both SMC and Best-of-N, we used $N = 32$ particles/samples. We sampled from the base model with temperature 0.8, and used systematic resampling in the implementation of SMC. We generated for a maximum of 3072 tokens for AIME and a maximum of 2048 tokens for Math500. Experiments were run on a mix of NVIDIA H100s and A100s.

As opposed to implementing SMC with $\mathcal{A}$ being the space of tokens (as in our experiments on prompt switching), we took $\mathcal{A}$ to be the space of *blocks* of some number B of tokens; doing so significantly improves the efficiency of SMC, as calls to the PRM are the main computational bottleneck. We set $B = 64$ for Math500 and $B = 128$ for AIME. Moreover, we observed that the reward model $\widehat{V}(a_{1:h})$ often gave unreliable answers when passed as input a partial completion $a_{1:h}$ which was truncated in the middle of a sentence or mathematical expression. Thus, we in fact truncated all blocks at the last instance of a delimiter character (e.g., new line, period).

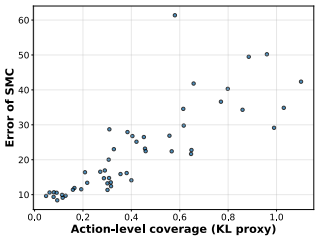


(a) AIME 2024 (b) AIME 2025

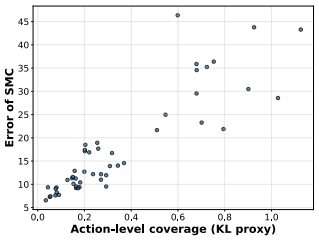
Figure 6: Performance of SMC vs. Best-of-N on AIME problems (each point is a different problem). Similar to Figure 2, the majority of points lie below the line $y = x$, indicating that SMC improves performance over Best-of-N on most problems.

H.8 Details for SMC on Math with Sharpening setup

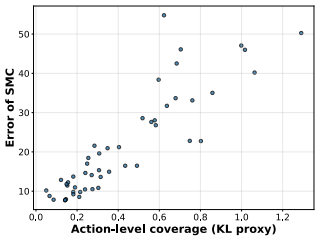
For the "sharpening" experiment discussed in Section 4.3 (Figure 4), we set $\pi_{\mathtt{ref}}$ to be Qwen2.5-7B-Instruct, and set $\widehat{V}(a_{1:h}) := \pi_{\mathtt{ref}}(a_{1:h})^{\lambda}$ for $\lambda \in \{1, 2\}$. We used $N \in \{1, 2, 4, 8, 16, 32\}$ particles/samples for both SMC and Best-of-N. We generated for a maximum of 1024 tokens, with block size of 64: namely, "actions" a_h correspond to blocks of 64 tokens, so that $H = 1024/64 = 16$. We used systematic resampling in the implementation of SMC, and adaptive resampling with a threshold of 0.5 (while theoretically analyzing such optimizations is not the focus of our work, they are commonly used in practice). Experiments were run on NVIDIA B200s.



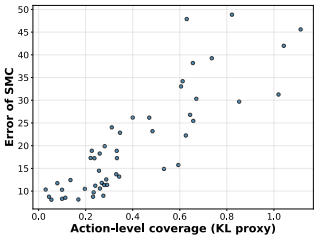
(a) $r = 0.83$



(b) $r = 0.89$

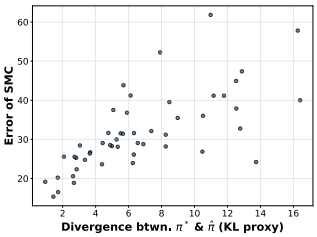


(c) $r = 0.88$

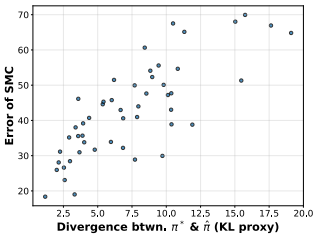


(d) $r = 0.86$

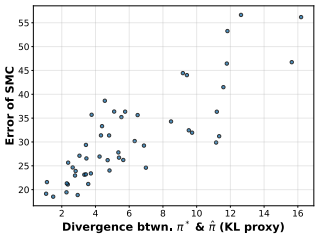
Figure 7: Additional instantiations of the experiment shown in Figure 1b, with different datasets (generated by GPT-5.2). Captions show Pearson correlation coefficients.



(a) $r = 0.70$



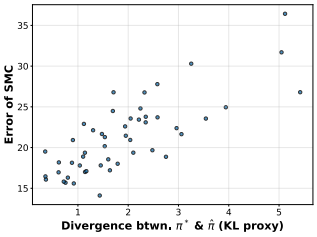
(b) $r = 0.79$



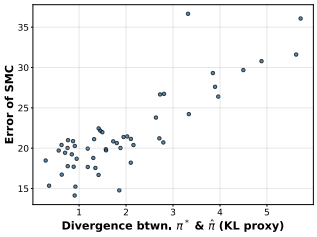
(c) $r = 0.83$

Figure 8: Additional instantiations of the experiment shown in Figure 1a, with different datasets (generated by GPT-5.2). Captions show Pearson correlation coefficients.

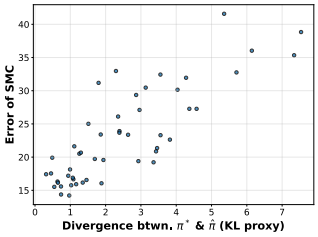
Finally, for the base distribution in the context of SMC, we sampled from $\pi_{\texttt{ref}}$ with temperature 0.8, meaning that, strictly speaking, using $\widehat{V}(a_{1:h}) = \pi_{\texttt{ref}}(a_{1:h})^{\lambda}$ would not in theory sample from the distribution $\pi^{\star}_{H}(a_{1:H}) \propto \pi_{\texttt{ref}}(a_{1:H})^{\lambda+1}$. The purpose of this experiment was not to closely follow all facets of the theoretical setup but rather explore how SMC performs in a natural, more complex, setting.



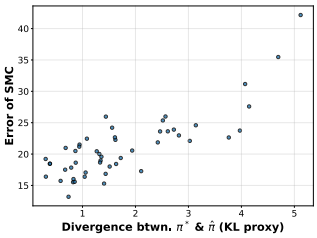
(a) $r = 0.78$



(b) $r = 0.85$



(c) $r = 0.83$



(d) $r = 0.82$

Figure 9: Additional instantiations of the experiment shown with the 4 datasets in Figs. 1a and 8, except with $\alpha = 1$. Captions show Pearson correlation coefficients.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes].

   Justification: The abstract and Section 1 state the paper's theoretical and empirical contributions, and the corresponding results are developed in Sections 3, 3.4 and 4. The discussion also identifies open directions and limitations of the current scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes].

   Justification: The paper discusses limitations in Sections 3.4 and 4.3 and in the Discussion (Appendix A), including horizon-dependence, gaps between sampling error and downstream accuracy, and the need for weaker or alternative notions of distributional closeness.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes].

Justification: The assumptions and theorem statements appear in Sections 2 and 3. Complete proofs are provided in Appendices D to G, with supporting tools in the technical appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes].

Justification: The experimental setup, models, metrics, horizons, particle counts, baselines, trial counts, and implementation choices are described in Section 4 and Appendices H.1 and H.5 to H.7.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No].

   Justification: The paper provides experimental details and cites the models and benchmarks used, which is sufficient information needed to reliably reproduce the results. We plan to release our code in the final version, but due to the structure and setup of our code and compute infrastructure doing so at present is not possible.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes].

   Justification: The paper specifies the evaluation settings, model choices, particle counts, horizons, temperatures, resampling choices, block sizes, trial counts, and metrics in Section 4 and the additional experimental details section in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes].

   Justification: The paper reports repeated-trial estimates for the Math500 and AIME comparisons, identifies the number of trials used, and reports Pearson correlation coefficients for the controlled prompt-switching experiments in Appendices H.5 to H.7.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes].

   Justification: The additional experimental details sections in the appendix report the type of GPU used for our various experiments. Figure 5 reports wall-clock scaling for the prompt switching experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes].

   Justification: The work consists of theoretical analysis and experiments with existing language models and standard reasoning/evaluation tasks. The paper does not involve human subjects, private data collection, or release of high-risk datasets or models.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A].

Justification: The paper's introduction discusses the motivation for the present work, namely that of improving inference capabilities of LLMs. Doing so in turn can have various positive and negative societal impacts, but due to the theoretical nature of the work, any direct societal impacts (positive or negative) are limited.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A].

Justification: The paper does not release new pretrained models, image generators, scraped datasets, or other assets that pose this kind of high-risk misuse concern.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes].

Justification: The paper credits and cites the existing models, benchmarks, and related methods used in the experiments, and respects all licenses and terms of use.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A].

Justification: The paper does not introduce or release a new dataset, code package, model, or benchmark asset as a contribution.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing experiments or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing or human-subject research, so IRB or equivalent approval is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes].

Justification: The authors used LLMs in the research process in now-standard ways, including for coding and assistance with low-level mathematical calculations. All proofs and computations were checked by the authors. The appendix also states that some prompt sets used for the controlled experiments were generated with GPT-5.2.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.